

SYNTHETIC BIOLOGY APPROACHES ESTABLISH THE FOUNDATION FOR
SUSTAINABLE PRODUCTION OF HIGH VALUE TERPENOIDS

By

Jacob David Bibik

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Cell and Molecular Biology—Doctor of Philosophy

2022

ABSTRACT

SYNTHETIC BIOLOGY APPROACHES ESTABLISH THE FOUNDATION FOR SUSTAINABLE PRODUCTION OF HIGH VALUE TERPENOIDS

By

Jacob David Bibik

Plants have become a promising platform for sustainable bioproduction of an array of natural products and specialty chemicals. Of particular interest are terpenes and the functionalized terpenoids, which represent the largest and most diverse class of natural products. These natural products are commonly used commercially as major constituents of flavorings and fragrances, oils, pigments, and pharmaceuticals, while having many other applications. Given the diversity and structural complexity of many terpenoids, they are often expensive and difficult, if not impossible, to chemically synthesize. Engineering these biosynthetic pathways in plant hosts may provide a sustainable platform to access terpenoids for industrial production. While plants offer a sustainable production platform, metabolic engineering for chemical production has largely focused on microbial hosts, and further development of strategies and tools for plant engineering is needed. In my dissertation, I have taken multi-pronged approaches to further develop sustainable bioproduction of terpenoids in plants. **First**, I developed strategies to optimize, re-target, and compartmentalize production of squalene, a C₃₀ triterpene, within plant cells to improve yields in plants. Re-targeting the final steps in squalene production, farnesyl diphosphate synthase (FDPS) and squalene synthase (SQS), from the cytosol to plastids enabled compartmentalization of biosynthesis away from competing cytosolic enzymes. I then anchored an optimized FDPS and SQS pair to the surface of cytosolic lipid droplets through fusions to the *Nannochloropsis oceanica* Lipid Droplet Surface Protein

(NoLDSP), where squalene can be sequestered and stored. Scaffolding the pathway to the surface of lipid droplets increased yields to more than twice that of plastidial targeting. Re-targeting this lipid droplet scaffolding to plastids, produced similar squalene yields as the soluble, plastid targeted pathway, and ameliorated some of the negative effects on photosynthesis. **Second**, I worked to engineer poplar, a bioenergy crop which emits large amounts of the hemiterpene isoprene, with these pathways as a platform for bioproduction and adding value to a bioenergy pipeline. Transformants were successfully created for plastid targeted squalene production, producing up to 0.63mg/gFW of squalene. The lipid droplet scaffolding strategies appeared toxic during tissue regeneration, suggesting a need for tissue specific engineering of these pathways in future iterations. **Third**, I developed a pipeline to identify, characterize, and engineer bidirectional promoters (BDPs), which enable divergent expression of two genes and improve gene stacking in plant constructs. As seen above with poplar, plant engineering is often limited by construct size, diverse promoter availability, and expression regulation, and a BDP library enables a range of expression in more compact constructs. I identified 34 BDPs from *Populus trichocarpa* and *Arabidopsis thaliana*, characterized their activity via *Nicotiana benthamiana* transient expression, and engineered select BDPs to further alter activities. Combining these BDPs with previously developed terminator sequences provided further regulation of expression. These genetic tools provide an array of expression activities and enable greater gene stacking options while offering the potential for more fine tuning of expression for multiple genes in a metabolic pathway. The work performed in this dissertation provide strategies to improve production of terpenoids in plants, establish production hosts, and engineer larger, complex pathways.

ACKNOWLEDGEMENTS

First, I would like to thank my research advisor, Dr. Björn Hamberger, for his amazing support throughout my studies. His guidance has meant a great deal to me, and he has become a true friend. Everyone within the Hamberger lab has been of great support as colleagues and friends, with a special thank you to Dr. Wajid Bhat who was instrumental in my scientific development. I would also like to thank my many collaborators who have provided their time and expertise in these projects, in particular Dr. Sarathi Weraduwege, Dr. Roberto Espinoza-Corral, and Dr. Faride Unda who were more than gracious to allow me to work closely with them. Dr. Margaret Petroff, Dr. Amy Ralston, and Alaina Burghardt were essential in the Cell and Molecular Biology program, and I greatly appreciate their support. My committee members Dr. Christoph Benning, Dr. Thomas Sharkey, Dr. Shawn Mansfield, and Dr. Daniel Ducat also provided great support through their collaborations and guidance, which has been key to my success. I have made many friends throughout my time at MSU and without them I may have withered away into an empty shell of a graduate student. My family has always supported my academic career and I am very thankful for them and especially Taylor Franklin, who has demonstrated great patience and encouragement as I have no doubt excessively vented about my many scientific roadblocks and frustrations throughout graduate school.

This research would not have been possible without the support of the U.S. Department of Energy, Great Lakes Bioenergy Research Center, under Award Numbers DE-SC0018409 and DE-FC02-07ER64494.

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
KEY TO ABBREVIATIONS	x
CHAPTER 1 Plant engineering to enable platforms for sustainable bioproduction of terpenoids	1
Abstract	2
Introduction	3
Terpenoid production in plants.....	4
Engineering strategies to improve bioproduction of terpenoids in plants	7
Compartmentalization of pathways.....	9
Engineering plants for production of terpenoids	10
Advancements in plant genetic engineering	12
Conclusions	16
Project goals and significance	17
CHAPTER 2 Pathway engineering, re-targeting, and synthetic scaffolding improves production of squalene in plants.....	19
Abstract	20
Introduction	21
Results and Discussion	25
Screening to improve the entry step in the MEP pathway	25
SQS and FDPS screening to improve squalene yields	28
Lipid droplet scaffolding optimization.....	30
Targeting LDSP scaffolds to plastids.....	33
Incorporating alternative contributions for the MVA pathway.....	37
Investigating how expression of squalene pathways affect photosynthesis	39
Methods.....	45
Genes synthesized and cloned	45
<i>Agrobacterium</i> -mediated transient expression, compound extraction, and measurement	46
Plastid fractionation and western blots	47
Gas-exchange measurements	49
Author contributions and acknowledgements	51
CHAPTER 3 Engineered poplar for bioproduction of the triterpene squalene ...	53
Abstract	54
Introduction	55
Results and Discussion	61
Engineering squalene production in transgenic NM6 poplar	61
Analysis of isoprene emission and photosynthesis	66

Technoeconomic analysis of poplar NM6 squalene production.....	69
Conclusions	71
Methods.....	71
Generation of poplar transformants.....	71
Transient expression in poplar NM6	73
Analysis of squalene production in transformants	73
Analysis of isoprene emission and photosynthesis	74
Technoeconomic analysis of squalene producing poplar	75
Author contributions and acknowledgements	76
CHAPTER 4 High-throughput identification, characterization, and engineering of plant bidirectional promoters	77
Abstract	78
Introduction	79
Results and Discussion	82
Identification and cloning of promoter sequences	82
Screening promoters in transient expression	84
Incorporating terminator sequences to further regulate expression.....	88
Engineering and characterizing synthetic BDPs.....	91
Conclusions	95
Methods.....	96
Promoter identification, selection, and cloning	96
<i>Agrobacterium</i> -mediated transient expression and measurement of fluorescent reporters	97
Author contributions and acknowledgements	98
CHAPTER 5 Conclusions and future directions.....	99
Summary	100
Future work.....	102
Developing compartmentalization of pathways	102
Developing poplar as a terpenoid production host	103
Engineering and expanding available plant BDPs.....	104
Conclusion.....	105
APPENDICES.....	107
APPENDIX A Supplemental Data for Chapter 2.....	108
APPENDIX B Supplemental Data for Chapter 3.....	115
APPENDIX C Supplemental Data for Chapter 4.....	120
REFERENCES	123

LIST OF TABLES

Table 2.1: Summary of changes in squalene yields and photosynthesis when scaffolding squalene biosynthesis using NoLDSP.	44
Table A.1: Genes used in Chapter 2 and their associated accession numbers.	109
Table A.2: Analysis of photosynthesis response to CO ₂ in leaves expressing plastid targeted and cytosolic squalene pathways, with and without NoLDSP scaffolding.	109
Table B.1: Economic parameters and assumptions used in technoeconomic analyses.	119
Table C.1: Promoters characterized in this study, the associated gene IDs, and reported native expression of each.....	121
Table C.2: Primers used to amplify candidate promoters from genomic DNA.....	122

LIST OF FIGURES

Figure 2.1: Overview of the engineering strategies developed or tested to improve squalene production in plants.....	22
Figure 2.2: Plastid targeted pathway optimization for squalene production.....	26
Figure 2.3: Screening of lipid droplet scaffolding combinations (a) and compatibility of alternative pathway contributions to engineered squalene pathways in the cytosol (b). 32	
Figure 2.4: Confocal microscopy of plastid targeted EYFP and EYFP-NoLDSP to compare membrane localization.	35
Figure 2.5: Western blots to determine membrane localization of plastid target, EYFP-NoLDSP fusion proteins.	36
Figure 2.6: Comparison of photosynthesis and squalene biosynthesis in leaves transiently expressing cytosolic and plastid targeted squalene pathways.	40
Figure 3.1: Representation of engineered squalene pathways used for poplar transformations.	59
Figure 3.2: Construct design for squalene pathways used for poplar transformations. .	62
Figure 3.3: Survey of transgenic poplar for production of squalene in leaf tissue.	64
Figure 3.4: Leaf stage, stem, and root analysis of squalene yields in select poplar transformants.	65
Figure 3.5: Analysis of isoprene emissions and photosynthesis in squalene producing poplar lines.....	68
Figure 3.6: Technoeconomic analysis of squalene production, extraction, and purification from poplar leaves.	70
Figure 4.1: Reported leaf expression of selected putative BDPs.	83
Figure 4.2: Reporter vectors developed for analysis of promoter activities.	85
Figure 4.3: Screening of putative BDPs identified from poplar and <i>Arabidopsis</i>	87
Figure 4.4: Testing select PtBDPs in combination with diverse terminator sequences. 90	

Figure 4.5: Prediction of TFBSs in PtBDP 3 and PtBDP16 to guide engineering design.	92
Figure 4.6: Synthetic BDPs created in this study.	93
Figure 4.7: Analysis of engineered promoter variants.	94
Figure A.1: Additional boxplots comparing soluble and NoLDSP scaffolding pathways in the cytosol and plastids.	110
Figure A.2: Confocal microscopy comparing cytosolic EYFP and EYFP-NoLDSP.	111
Figure A.3: Additional plastid fractionation and western blots demonstrating EYFP- NoLDSP localization in chloroplast membranes.....	112
Figure A.4: Comparison of the effects of transient expression of plastid targeted squalene pathways in leaves and subsequent effects on photosynthesis compared to controls.....	113
Figure A.5: A/Ci curves comparing plastid targeted and cytosolic squalene pathways, with and without NoLDSP scaffolding.....	114
Figure B.1: Transient expression of lipid droplet scaffolding in poplar NM6 leaves.....	116
Figure B.2: Additional analysis of isoprene emission and photosynthesis in squalene producing poplar NM6.	117
Figure B.3: Parameters used to the technoeconomic analyses performed for squalene extraction and purification from poplar leaves.	118

KEY TO ABBREVIATIONS

MVA	Mevalonate
MEP	Methylerythritol 4-phosphate
IDP	Isopentenyl diphosphate
DMADP	Dimethylallyl diphosphate
FDP	Farnesyl diphosphate
GDP	Geranyl diphosphate
GGDP	Geranylgeranyl diphosphate
SQS	Squalene synthase
FDPS	Farnesyl diphosphate synthase
LDSP	Lipid Droplet Surface Protein
GGDPS	Geranylgeranyl diphosphate synthase
DXS	1-deoxy-D-xylulose-5-phosphate synthase
nDXS	Novel DXS-like
HMGR	3-hydroxy-3-methylglutaryl-CoA reductase
GAP	Glyceraldehyde 3-phosphate
Ru5P	Ribulose 5-phosphate
CasS	Casbene synthase
IDI	Isopentenyl diphosphate isomerase
WRI1	WRINKLED1
MPD	Phosphomevalonate decarboxylase
IPK	Isopentenyl phosphate kinase

BCCP1	Biotin carboxyl carrier protein 1
SBPase	Sedoheptulose-1,7-bisphosphatase
FBN1a	Fibrillin 1a
PsaF	Photosystem I subunit F
TOC75	Translocon of the Outer Chloroplast
C_i	Intercellular CO ₂ concentration
A	Photosynthetic rate
V_{cmax}	Rubisco activity
J	Electron transport rate
Φ_{PSII}	Operational efficiency of photosystem II
g_{sw}	Stomatal conductance
TPU	Triose phosphate utilization rate
GC-FID	Gas chromatography with flame ionization detection
CaMV	Cauliflower Mosaic Virus
NM6	<i>Populus nigra</i> L. x <i>Populus maximowiczii</i> A. Henry
P39	<i>Populus alba</i> x <i>Populus grandidentata</i>
MCS	Multiple cloning site
gFW	Grams fresh weight
ISPS	Isoprene synthase
TEA	Technoeconomic analysis
EYFP	Enhance yellow fluorescent protein
RFP	Red fluorescent protein
GFP	Green fluorescent protein

BDP	Bidirectional promoter
sBDP	Synthetic bidirectional promoter
UDP	Unidirectional promoter
TF	Transcription factor
TFBS	Transcription factor binding site
UTR	Untranslated region
NosT	Nopaline synthase terminator
MasT	Mannopine synthase terminator
BDB501	Bean dwarf mosaic virus movement protein 3' end
AtHSP	<i>Arabidopsis thaliana</i> heat shock protein terminator
NbHSP	<i>Nicotiana benthamiana</i> heat shock protein terminator
Rb7 MAR	Rb7 matrix attachment region
MSP	Minimum selling price

CHAPTER 1

Plant engineering to enable platforms for sustainable bioproduction of terpenoids

Abstract

Terpenoids make up the most diverse class of natural products, a number of which currently have important biotechnological roles. Many terpenoids are difficult or impossible to chemically synthesize, so there is great interest in developing sustainable bioproduction platforms to recreate these pathways in host organisms. In addition to containing both pathways that generate the terpenoid building blocks as well as the cell structures and compartments required for many of the enzymes involved, plants may provide a sustainable, low input system to produce these chemicals. There have been many recent advancements in discovery of pathways to terpenoids of interest as well as strategies to engineer commercially relevant yields in host plants. While there are many researchers working to discover the biosynthetic pathways and gain access to production of novel terpenoids, I will mostly focus on advancements towards engineering plants for production of terpenoids. I will highlight strategies currently used to produce target products, optimization of known pathways to improve yields, compartmentalization of pathways within cells, and genetic tools developed to facilitate complex engineering of biosynthetic pathways. These advancements have enabled the use of plants as hosts for bioproduction of terpenoids.

Introduction

Plants contain enormous chemical diversity, which humans have been utilizing for thousands of years, especially from traditional medicinal plants^{1,2}. Much of this diversity has exploded as plants have developed an array of specialized metabolites, many of which have emerged in response to various environmental conditions, both biotic and abiotic, and for signaling and communication. An estimated number of more than one million unique metabolites produced across all plants³, with over 200,000 predicted to be specialized metabolites⁴, establishes plants as natural chemical factories that can be further engineered for biotechnological applications^{5,6}. Terpenes, and the further functionalized terpenoids, are a particularly diverse group of natural products with more than 50,000 of the over 60,000 reported structures having been identified in plants, according to a recent analysis of the Dictionary of Natural Products⁷. While terpenoids have important roles within central metabolism across plant species, the majority have evolved as specialized metabolites, often only found in specific plant lineages or species^{8–10}. They have evolved to become the largest class of natural products and filled a broad spectrum of roles within plants, but the chemical diversity of terpenes and terpenoids far surpass their known roles in plants. As a consequence of this expansive diversity, these chemicals have been exploited throughout human history as major components of many herbs and spices, medicinal plants, food crops, resins, and many other traditional plants¹¹.

Recently, there has been growing interest in taking synthetic biology approaches to engineer plants into chemical factories through development of biosynthetic pathways for terpenoids important to modern society, while even further expanding beyond naturally occurring terpenoid chemistry^{8,10,12–15}. Discovery and engineering of these biosynthetic

pathways has enabled production of not only known terpenoids, but the creation of novel structures through combining modular pathways^{16,17}. Much of the success in plant engineering has involved more simple terpene pathways, but strategies are being developed to build more complex pathways and engineer how, where, and when terpenoids are produced within plants. In addition to the innovation seen with pathway engineering, there have been advancements in genetic engineering tools, yet more advancement is needed to enable engineering of longer and more complex pathways to include not only the terpene synthases, but downstream enzymes responsible for functionalization of terpene cores. Here I will review efforts made to engineer plants for production of terpenoids, engineering and optimization of biosynthetic pathways, compartmentalization of pathways and products, and advancements in genetic tools to enable engineering of plant hosts with larger and more complex pathways.

Terpenoid production in plants

Terpenoids are universally found across all kingdoms of life and are formed from common C₅ building blocks. These building blocks, isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMADP), are isomers synthesized via two pathways that are both present in plants, though most organisms contain only one. The mevalonate (MVA) pathway, which is commonly found in eukaryotic organisms, synthesizes a cytosolic pool of IDP/DMADP in plants starting with condensation of three acetyl-CoA molecules in two enzymatic steps to form 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA). HMG-CoA is then reduced by HMG-CoA Reductase (HMGR) to form mevalonate, the dedicated step to the MVA pathway. The methylerythritol 4-phosphate (MEP) pathway, often found in prokaryotic organisms, synthesizes a plastidial pool of IDP/DMADP in plants, which

begins with condensation of pyruvate and glyceraldehyde 3-phosphate (GAP). This condensation forms 1-deoxy-d-xylulose-5-phosphate (DXP) and is catalyzed by DXP synthase (DXS). In addition to this natural compartmentalization of IDP/DMADP accumulation, the enzymes involved in terpenoid biosynthesis are also co-localized to access the building blocks in each compartment. Typically localized to plastids are the diphosphate synthases which form the C₁₀ geranyl diphosphate (GDP) or the C₂₀ geranylgeranyl diphosphate (GGDP) through condensation of two or four IDP/DMADP molecules by GDP synthase or GGDP synthase (GGDPS), respectively. Furthermore, mono- and di- terpene synthases co-localized to plastids utilize GDP and GGDP to synthesize the C₁₀ monoterpenes and C₂₀ diterpenes, respectively. Some plants are native emitters of the C₅ hemiterpene isoprene and contain an isoprene synthase (ISPS), which is also typically plastid localized. Cytosolically, three IDP/DMADP molecules are formed into the C₁₅ intermediate farnesyl diphosphate (FDP) through condensation by FDP synthase (FDPS), which serves as the substrate for sesquiterpene synthases to synthesize sesquiterpenes. Localized to the endoplasmic reticulum is squalene synthase (SQS), which condenses two FDP molecules to form the C₃₀ squalene, which is the precursor to sterols and triterpenoids.

There have been many research efforts to understand and engineer these pathways while developing hosts for biosynthesis of terpenoids as direct products and intermediates in semi-chemical synthesis. In some instances, plant species that are known to synthesize particular terpenoids have been manipulated to enable larger scale production. For example, the diterpene (-)-sclareol from *Salvia sclarea* (clary sage) is commonly extracted as an intermediate of the semi-chemical synthesis of ambroxide, a

high-value product used in the fragrance industry¹⁸. Ambroxide was historically extracted from sperm whale ambergris but the biosynthetic pathway is unknown, though bacterial pathways have recently been engineered to synthesize the triterpenoid precursor, ambrein, of which ambroxide is an oxidized product of^{19,20}. In another example, plant tissue culture techniques were developed as the production platform for paclitaxel, a complex diterpenoid made in yew trees (*Taxus* spp.)²¹. Under standard growth conditions paclitaxel was not produced at industrially relevant yields, but development and optimization of tissue culture enabled production for the compound to become widely used in chemotherapeutics. While these examples demonstrate ingenuity to produce high value terpenoids from native species, this is not feasible for most plants or terpenoids. To this end, engineering plants with specific biosynthetic pathways and strategies to optimize yields has become a focus in synthetic biology.

While microbial fermentation is often used for production of chemicals like terpenoids, engineered plants present an opportunity for a low input chassis system by utilizing sunlight and CO₂ to produce large amounts of biomass^{22,23}. This can be especially effective in high biomass producing, non-food crops, which can be grown on marginal lands not suitable for food crops. Additionally, plants already contain the cellular structures, co-enzymes, and precursor pathways to support production of many natural products. There are many examples of effective plant engineering to produce valuable terpenoids with biotechnological importance. Many terpenoid pathways have been engineered in tobacco (*Nicotiana* spp.), which has become extensively used due to ease of engineering transgenic plants, as well as transient expression to rapidly test pathways^{8,12–15,17,24–28}. Other model plants have also been used for terpenoid production,

including tomato²⁹, Arabidopsis^{30,31}, and even moss^{32,33}. These plants have been used to synthesize an array of terpenoid classes, many of which may have significance for biotechnological applications.

These platforms have been employed to produce volatile terpenoids including the hemiterpene isoprene²⁵, monoterpenoids^{12,15,30,31}, and sesquiterpenoids^{12,15,30–33}, many of which are commonly used in the fragrance industry. The longer chain, non-volatile terpenoids have also been of great interest, including strategies to expand the number of diterpenoids that can be produced *in planta*^{8,17} and have potential roles in a range of industries. Additionally, there have been efforts to engineer the production of squalene as well as downstream triterpenoids and sterols, which have applications in cosmetic oils, biofuels, or pharmaceuticals^{13,14,27}. Finally, there has been a focus in engineering tetraterpenoids and, in particular, carotenoids which are of interest as natural pigments and nutritional additives^{24,28}. While many studies have enabled synthesis of an array of terpenoids, they often result in low yields because they are not optimized for production. I will next discuss strategies being developed to improve production in plant hosts, through pathway optimization compartmentalization.

Engineering strategies to improve bioproduction of terpenoids in plants

Approaches have been developed to increase metabolic flux towards desired terpenoids including overexpression of key bottlenecks in the MEP/MVA pathways, engineering de-regulated variants of enzymes^{34,35}, incorporation of alternative contributors to influence IDP/DMADP pools^{30,31,36}, and silencing of competing pathways²⁶. Each of these approaches either improve availability of, or redirect IDP/DAMDP for production of desired terpenoids. For example, overexpression of rate

limiting steps of the MVA or MEP pathways, HMGR and DXS, enable significant increases to terpenoid by increasing metabolic flux towards IDP/DMADP^{36–39}. These enzymes have also been targeted for engineering to reduce feedback inhibition from the end products IDP and DMADP^{34,35}. Other strategies have been developed via alternative contributions to the MVA/MEP pathways or to IDP/DMADP formation directly. One set of enzymes, phosphomevalonate decarboxylase (MPD) and isopentenyl phosphate kinase (IPK), were used to recreate an archaeal pathway in plants, which performs the final two steps of the MVA pathway to form IDP/DMADP, but essentially in reverse order^{30,31}. Another group found overexpressing the gene for a Biotin Carboxyl Carrier Protein 1 (BCCP1) can disrupt proper formation of the acetyl-CoA carboxylase complex, increasing availability of acetyl-CoA to enter the MVA pathway by reducing conversion to malonyl-CoA in plastids³⁶. Virus-induced gene silencing has been pursued to reduce conversion of shared precursors and redirect them towards desired terpenoids. For example, silencing expression of phytoene synthase which converts GGDP to phytoene, the carotenoid precursor, increased production of the diterpene taxadiene which is also derived from GGDP²⁶. There are also other enzyme engineering strategies from microbes that may be of value in plants. For example, bacterial enzyme variants have been developed to also synthesize DXP, but from ribulose 5-phosphate, potentially providing an unregulated mechanism to improve MEP pathway flux in plants⁴⁰. Additionally, random mutagenesis approaches developed in *Escherichia coli* to improve enzyme functions may be a strategy to rapidly improve enzymes which can be utilized for plant production. Engineered GGDP variants created through random mutagenesis showed improved lycopene

production in screening, as well as increased production of the diterpene levopimaradiene⁴¹.

Compartmentalization of pathways

Strategies have been developed to improve terpenoid production by re-targeting pathways between the cytosol and plastids^{12–15,28,42}, create synthetic compartments^{14,15}, and expression of pathways to specific tissues^{43–49}. Plants naturally compartmentalize terpenoid biosynthesis within cells as well as across different tissue types, which can be utilized to engineer novel pathways. Re-targeting terpenoid biosynthesis from the cytosol to plastids has been shown to improve terpenoid production^{12,13,42}, as has re-targeting natively plastidial pathways to the cytosol²⁸. Mitochondria have also been targeted for production of terpenoids, in particular sesquiterpenes which are typically produced cytosolically^{42,50–52}. Hijacking native organelles may enable production of a target compound while reducing negative regulation or competition present in the native compartments. Furthermore, strategies have been developed to improve production of hydrophobic terpenoids using lipid droplets as synthetic storage compartments^{14,15}. Re-engineering subcellular compartmentalization can not only improve terpenoid production, but also alleviate negative effects these pathways may have on the plant host^{13,14}. Manipulation of compartments and subcellar localization is therefore an effective strategy to further engineer improved terpenoid production in plants.

In addition to the natural subcellar localization of different pathways, plants have naturally developed differential expression and accumulation of pathways and products across different tissues which also provide opportunities for engineering^{10,53–57}. Terpenoid biosynthesis is seen in specific tissues like roots, leaves, and flowers, and accumulation

of terpenoids can even be localized to specific cell types and structures associated with these tissues. For example, specialized cork cells⁵³ or secretory ducts⁵⁵ in root tissue, resin ducts in stems⁵⁸, and other oleoresin structures commonly found in conifers⁵⁹ have been shown to synthesize and store terpenoids. Similarly, glandular trichomes on leaf tissue are known to accumulate a variety of specialized metabolites including terpenoids and can be specifically engineered for terpenoid production⁶⁰. Flowers are also known to specifically produce terpenoids, especially volatile variants⁵⁴, and can even emit and store these products in the stigma through a natural fumigation process in unopened flower buds⁵⁶. Building from natural biosynthesis examples, engineering specific tissues, cell types, and structures may allow for greater control of terpenoid production in plant hosts.

Engineering plants for production of terpenoids

Development of engineering strategies is often performed using *Agrobacterium*-mediated transient expression, which allows rapid testing of pathways and biological parts in a matter of days as opposed to the months or years that are often required when creating stable transgenic plants. Pathways can be quickly tested in a combinatorial approach by mixing parts which enables characterization of biosynthetic pathways and construction of novel terpenoid pathways^{8,17,27}. Transient expression is typically localized to the infiltrated leaf tissue, though some methods have been developed to transiently express in other tissues^{61,62}. It is possible to produce potentially commercially relevant yields of terpenoids using *Agrobacterium*-mediated transient expression in lab plants like tobacco²⁷, but stable transformation may allow more sustainable and large scale production of compounds. Furthermore, engineering terpenoid production in crops with

industrial uses may be a strategy to add value to existing infrastructure while reducing the cost to produce the terpenoids of interest⁶³.

Engineering fast growing, high biomass producing crops like poplar and sorghum (*Sorghum* spp.) has become of great focus to generate bioenergy and bioproduct feedstocks. These crops are typically desirable as they represent significant lignocellulosic feedstocks that can be converted to simple sugars and monolignols for microbial production of biofuels and bioproducts, like terpenoids^{64,65}. Their robust growth and established transformation protocols also make bioenergy crops a candidate for engineering production of specialty chemicals and other bioproducts. A technoeconomic analysis modeling sorghum biomass showed that engineering the crops to produce a variety of compounds, including terpenoids, may improve the economics when extracting the compounds prior to lignocellulosic biomass processing⁶³. This strategy may also be effective for a woody bioenergy crop like poplar, which is also commonly used in the pulp and paper industry and the manufacture of oriented strand board. Poplar has previously been engineered for production of specialty chemicals derived from aromatics that are also used in biosynthesis of the monolignols which make up the lignin biomass^{66,67}. These studies demonstrate poplar is not only valuable as a bioenergy crop, but also as a platform for direct production of high value chemicals. Poplar has also been extensively studied because many species have the ability to synthesize and emit substantial amounts of isoprene, suggesting the metabolic capacity to produce large amounts of terpenoids if engineered⁶⁸. Furthermore, isoprene has a large role in climate change, as massive amounts, over 500 Tg year⁻¹ from all plants⁶⁹, are emitted each year and reducing emissions in commercial poplar may be a strategy to help combat climate change⁷⁰. To

this end, engineering poplar to re-route IDP/DMADP away from isoprene biosynthesis and towards production of target terpenoids may be effective in producing high value products while reducing isoprene emission in poplar plantations.

Combinatorial assays and co-expression of complex pathways are easily performed using transient expression by mixing of *Agrobacterium* strains harboring different plasmids with target genes inserted, even if in the same vector background. Transient expression does not require gene stacking, nor does it rely on plant selection markers because genomic integration is not required. Additionally, transient expression typically relies on strong constitutive promoters, like the cauliflower mosaic virus (CaMV) 35S promoter⁷¹, when attempting to produce large amounts of products, meaning there is little concern over temporal or spatial expression regulation. These expression constructs can also be paired with over-expression of the gene encoding viral P19 suppressor protein which can suppress RNA silencing from the host plants, as is seen in the *pEAQ* vector series⁷². Including viral silencing suppressors like P19 when generating stable transformants, however, often leads to developmental issues as the suppression of RNA silencing is not specific⁷³. Therefore, many of the strategies that have been optimized for transient expression are not easily translatable to generating stable transformants, but recent innovations in genetic engineering are enabling more complex pathway design.

Advancements in plant genetic engineering

When engineering larger and more complex pathways in stable transformants, the plants are typically transformed with multiple genes through one of three strategies: (i) consecutive re-transformations with different genes, (ii) co-transformations with genes on

multiple constructs, or (iii) transformation with multigene constructs^{74,75}. Consecutive re-transforming can take months or years, co-transformations are inefficient and require multiple selection markers, and multigene platforms require several unique promoters, resulting in large constructs with low transformation efficiency. There have been advancements in the assembly of larger constructs which enable more efficient transformation with large DNA fragments. For example, Collier *et al.* 2018 developed a recombination system to assemble a large construct with a 28.5kb transfer DNA region, which required multiple promoters and two selection markers to ensure genomic integration and expression of all genes⁷⁶. Technologies like this provide strategies to begin building larger constructs, but diverse genetic elements are still needed for more complex metabolic engineering in plants.

There are a set of genetic tools that have been traditionally used to reliably overexpress genes of engineered pathways. These are from viral sources like the CaMV 35S promoter⁷¹ and terminator⁷⁷, bacterial sources like the promoter and terminator sequences from various opine synthases⁷⁸, or plant sources like the maize Ubiquitin promoter⁷⁹ or the rice Actin promoter⁸⁰. Additionally, the use of 5' and 3' untranslated regions (UTRs) have been applied to increase expression, in particular the UTRs from cowpea mosaic virus⁸¹ which are implemented in the pEAQ-*HT* vectors⁷². These regulatory tools have proven effective within plant biotechnology when engineering high expression of genes and pathways, but additional tools are being developed to further improve expression regulation, tissue specificity, and gene stacking abilities from natural and synthetic elements^{43–49,82–96}.

One of the most important considerations for metabolic engineering is regulation of gene expression strengths and tissue specificity, which is largely controlled by promoters in plants. A central theme has been on creating synthetic promoters for robust expression, mainly through combining known elements from different sequences, which has been recently reviewed⁸². In general, synthetic promoter design in plants has not been as advanced as microbial systems, but there have been major strides recently in developing synthetic and tunable plant promoters^{84–88}. For example, Cai *et al.* 2020 developed a strategy to computationally design constitutive, synthetic minimal promoters which demonstrated a range of expression strengths⁸⁵. In another recent study by Jores *et al.* 2021, a comprehensive analysis of *Arabidopsis*, *Zea mays*, and *Sorghum bicolor* core promoter sequences was performed and used to also create a series of synthetic variants⁸⁷. Another recent review has summarized advancements in synthetic regulation of pathways through post-transcriptional and translational approaches in addition to promoter design⁸³. Post-transcriptional engineering approaches include UTRs on the 5' and 3' ends of a transcript^{89,90}, as well as terminator sequences⁹¹. Synthetic riboswitches have also been developed for translational control of pathways^{97,98}, enabling inducible translation of mRNA from target genes. These strategies enable additional layers of regulation through influencing gene expression, mRNA stability, and translation^{99,100}.

In addition to expression tunability, regulating where multigene pathways are expressed may be especially important to dictate where products like terpenoids accumulate and to reduce potential adverse effects of this accumulation on plant development. To this end, a number of promoters have been developed to regulate tissue specificity^{43–49}. Engineering high specificity can prove advantageous as has been

demonstrated in leaf oil production using a leaf senescence specific promoter to reduce pleiotropic effects of accumulation while obtaining oilseed-like levels in more biomass than traditional seed production⁹². In addition to tissue specificity, strategies have been developed to efficiently express multiple genes in a single construct. For example, one system has been developed where a synthetic activator gene is placed under the control of an endosperm specific promoter, which when expressed activates expression of multiple downstream genes under the control of synthetic promoters responsive to the synthetic activator⁸⁴. A simpler approach to expression of multiple genes is the use of bidirectional promoters, which have been isolated from native genomic sequences^{45,46,93,94} or created synthetically through stacking unidirectional promoters, like the CaMV 35S, head-to-head^{49,95,96}. Bidirectional promoters have been well studied in microbial engineering^{101,102}, but have yet to be used more broadly for plant engineering. In combination with linker peptides like the self-cleaving 2A peptide from the foot-and-mouth disease virus¹⁰³ or the more efficient hybrid LP4/2A liker^{103,104}, bidirectional promoters may enable polycistronic expression on either side of the promoter. These strategies to regulate expression of multiple genes would enable more complex gene stacking to build entire metabolic pathways in a single construct used for plant transformation.

There have been many advancements in genetic tools for plant engineering and these will aid in engineering more precise regulation of metabolic pathways. Combining tunable expression regulation, tissue specificity, and compact construct assemblies will enable complex engineering of large multigene metabolic pathways. While many of these tools only show functionality within specific plant species, they may be more broadly

applicable, or at least provide inspiration to further engineer regulatory tools for chassis species more predisposed for production of bioproducts like terpenoids. Additionally, improving gene stacking strategies will improve engineering of plants to avoid issues like repeating promoter sequences being silenced, limitations in selectable markers, and smaller constructs for more reliable transformation with a larger number of genes.

Conclusions

While plants are natural chemical factories, advancements in metabolic engineering have enabled redesigning plants as chassis for more economically viable production of chemicals like terpenoids. With developments in engineering the MEP and MVA pathways and downstream enzymes for terpenoid production industrially relevant yields are becoming more accessible. To further push yields to become economically viable, it will be important to consider the optimization of pathways, compartmentalization of pathways and storage of products, and the proper plant host for production. Furthermore, developing the genetic tools used for construction of pathways in hosts will be key. Developing tissue and cell type specificity, along with intracellular compartmentalization, would enable precision engineering of terpenoid production which could improve yields while reducing negative effects accumulation of products has on the host. Creative gene stacking with diverse promoters will enable expansion to engineering larger terpenoid pathways while reducing potential silencing due to expression from repeated promoter sequences. Furthermore, expansion of promoter libraries with varying expression strengths will allow finer expression tunability for each gene in a pathway, providing the tools to begin regulating metabolic pathway stoichiometry in plants.

Project goals and significance

Despite recent advancements in terpenoid metabolic engineering in plants, further development is needed of strategies to produce commercially relevant yields and genetic tools to facilitate engineering of complex pathways. Many studies have focused on strategies to increase IDP/DAMDP supply and flux towards terpenoid production^{30,31,35–40}, but another area of important consideration is terpenoid storage within hosts^{14,15}. Additionally, engineering terpenoid production in plants requires not only installation of the desired pathway, but also the enzymes involved in boosting IDP/DAMDP supply and establishing storage compartments. To enable engineering of these complex pathways, development of gene stacking strategies and regulatory tools are needed⁷⁶. Therefore, my dissertation projects aimed to develop strategies which improve yields and production capacity of terpenoids in plants while also developing novel genetic tools to facilitate the engineering of complex pathways.

In this dissertation, I present synthetic biology approaches to redesign plants as hosts for production of terpenoids through a multipronged effort. In Chapter 2, I re-engineer plastids and lipid droplets as synthetic storage compartments in a transient expression system and determine storage capacity of products to be a significant limitation for improving yields. These approaches also demonstrate the ability to ameliorate negative effects terpenoid accumulation has on hosts, providing a strategy to improve the overall engineered system. In Chapter 3, I implement these strategies in a stable transgenic production host, moving beyond lab scale proof of concept. Transgenic lines demonstrated promising terpenoid yields which lay the groundwork to make such a platform economically viable. Furthermore, findings here highlight the need for greater

regulation of these pathways, in which tissue specificity may be particularly important. In Chapter 4, I characterized a library of BDPs which showed a range of expression strengths in a heterologous system, demonstrating functionality as genetic tools for broader host species. These BDPs were reported to have a range of leaf specificity in their native species, which may provide additional control over engineered pathway regulation. This work also establishes a pipeline to identify, characterize, and engineer BDPs which enables rapid and high-throughput development of promoter libraries. Together, the research presented in this dissertation contributes multiple approaches to not only improve the strategies for production of terpenoids, but also the tools to facilitate more efficient engineering of complex, multi-gene pathways in plant systems.

CHAPTER 2

Pathway engineering, re-targeting, and synthetic scaffolding improves production of squalene in plants

Results from this chapter have been adapted with permission from Bibik, J.D., Weraduwaage, S.M., Banerjee, A., Robertson, K., Espinoza-Corral, R., Sharkey, T.D., Lundquist, P.K., and Hamberger, B.R. (2022). Pathway Engineering, Re-targeting, and Synthetic Scaffolding Improve the Production of Squalene in Plants. *ACS Synth. Biol.* <https://doi.org/10.1021/acssynbio.2c00051>. Copyright 2022 American Chemical Society.

Abstract

Plants are increasingly becoming an option for sustainable bioproduction of chemicals and complex molecules like terpenoids. The triterpene squalene has a variety of biotechnological uses and is the precursor to a diverse array of triterpenoids, but we currently lack a sustainable strategy to produce large quantities for industrial applications. Here, I further establish engineered plants as a platform for production of squalene through pathway re-targeting and membrane scaffolding. The squalene biosynthetic pathway, which natively resides in the cytosol and endoplasmic reticulum, was re-targeted to plastids, where screening of diverse variants of enzymes at key steps improved squalene yields. The highest yielding enzymes were used to create biosynthetic scaffolds on co-engineered, cytosolic lipid droplets, resulting in squalene yields up to 0.58 mg/gFW, or 318% higher than a cytosolic pathway without scaffolding during transient expression. These scaffolds were also re-targeted to plastids where they associated with membranes throughout, including formation of plastoglobules, or plastidial lipid droplets. Plastid scaffolding ameliorated negative effects of squalene biosynthesis and showed up to 345% higher rates of photosynthesis than without scaffolding. This study establishes a platform for engineering production of squalene in plants, providing the opportunity to expand future work into production of higher-value triterpenoids.

Introduction

Engineered plants present an opportunity for sustainable production of high-value chemicals important for many industries. One class of chemicals with growing interest are terpenoids, the most diverse class of natural products with an array of biotechnological applications. The C₃₀ triterpene squalene is a long-chain hydrocarbon and the precursor to sterols and triterpenoids¹⁰⁵. Since first being described in shark liver oil in 1916¹⁰⁶, it has been developed for a number of commercial uses that include cosmetic oils, as a vaccine adjuvant, and has potential as an energy dense biofuel^{107,108}. Squalene is also an important intermediate in the production of higher-value derivatives, such as the triterpenoid ambrein and its derivative (-)-ambrox, which are used in the fragrance industry^{109,19,20}. For commercial applications, squalene has historically been obtained from shark liver and more recently vegetable oils¹¹⁰, but engineered crops may be able to produce it with higher specificity and yields. Establishing more sustainable plant production strategies for squalene and the derived triterpenoids may enable an economically viable platform for supply to a range of industries. It has also been suggested that incorporating engineered biosynthetic pathways into bioenergy crops may improve the financial feasibility of both terpenoid production and conversion of plant biomass to biofuels⁶³. Using plants to produce squalene and valuable derivatives requires innovation to increase yields while reducing potentially negative effects of engineered pathways on the host^{24,111,112}.

The terpene backbones from which terpenoids are derived are assembled in five carbon segments through condensation of the building blocks isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMADP)¹¹³. In plants, these building blocks are

synthesized either in the cytosol through the mevalonate (MVA) pathway, starting with condensation of 2 acetyl-CoA molecules, or plastids through the methylerythritol 4-phosphate (MEP) pathway, starting with pyruvate and glyceraldehyde 3-phosphate (GAP) condensation (Figure 2.1). Also localized to the cytosol is farnesyl diphosphate synthase (FDPS) which catalyzes the head-to-tail condensation of one DMADP and two IDP molecules to form the C15 farnesyl diphosphate (FDP). Two of these FDP molecules are then condensed head-to-head by the endoplasmic-reticulum-bound squalene synthase (SQS) to form squalene. Multiple approaches have been taken to increase terpenoid yields in plants through engineering or re-targeting of these pathways.

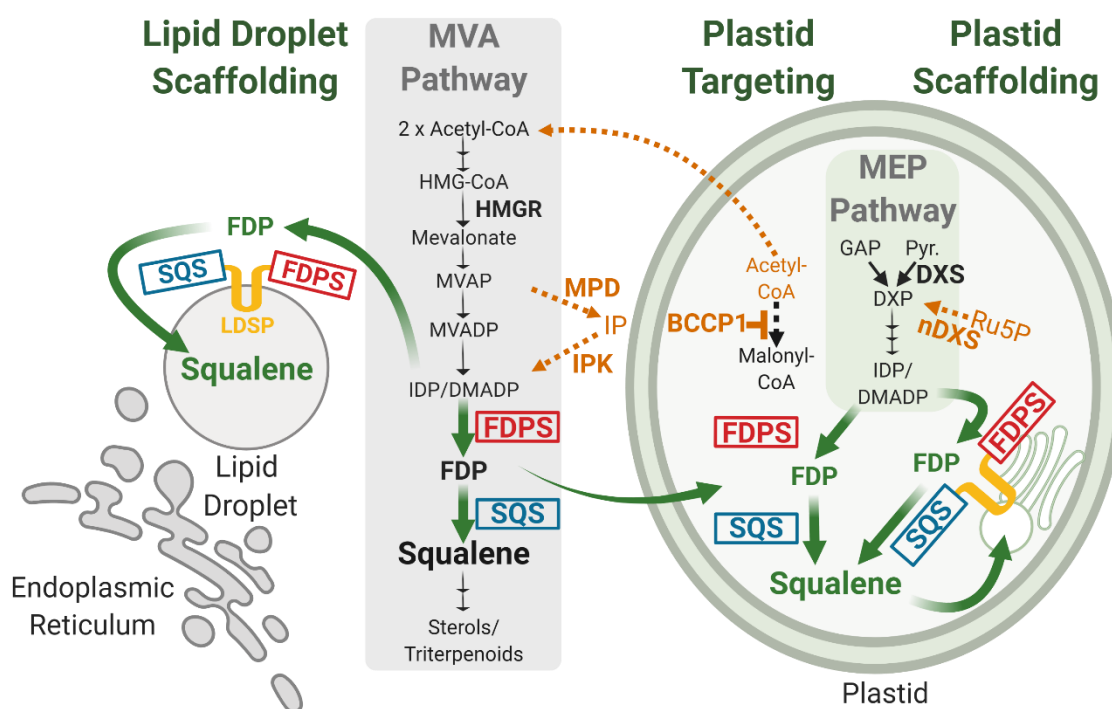


Figure 2.1: Overview of the engineering strategies developed or tested to improve squalene production in plants. Green arrows indicate the engineered squalene pathways developed in this study, scaffolding pathways on cytosolic lipid droplets (left), or re-targeting to plastids (right) with and without scaffolding. Orange pathways indicate alternative contribution strategies tested, where genes were co-expressed with genes for various squalene pathways.

Both the MVA and MEP pathways are regulated through multiple mechanisms, many of which have become targets for engineering^{113–118}. Common strategies to overcome regulatory limitations include overexpression and engineering of key enzymes in both pathways. Overexpression of the genes for 1-deoxy-d-xylulose-5-phosphate synthase (DXS), the first step in the MEP pathway, and 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR), the committed step to the MVA pathway (Figure 2.1), have been shown to generate abundant supply of terpenoid precursors IDP and DMADP, while increasing terpenoid yields^{37,39,38,36}. Both DXS and HMGR have been targets for engineering, where variants have been created and shown to have reduced negative regulation in plant systems^{34,35}. Other studies have indicated alternative contributions to the precursor pathways or IDP and DMADP pools can also improve terpenoid production^{30,31,36}. Overexpression of the *Arabidopsis thaliana* biotin carboxyl carrier protein 1 gene (*AtBCCP1*) was shown to improve acetyl-CoA availability and utilization by the MVA pathway to increase terpenoid yields³⁶, by potentially disrupting the acetyl-CoA carboxylase complex. Addition of a phosphomevalonate decarboxylase from *Roseiflexus castenholzii* (*RcMPD*) and an *Arabidopsis* isopentenyl phosphate kinase (*AtIPK*), a non-canonical route to IDP using MVA pathway intermediates, was also found to improve terpenoid production in plants^{30,31}. These studies provide potential biological parts for optimization or combinatorial approaches to further develop plant systems for terpenoid production

Re-targeting and compartmentalization of terpenoid pathways has enabled storage of products to increase yields^{12,15,42} and reduce negative effects of product accumulation on plants^{13,14}. Previous work has shown that redirecting terpene

biosynthesis from the cytosol to plastids in plants, using the organelle as a storage compartment, can increase yields^{12,13,42}. In other work, lipid droplets have been adapted as synthetic storage compartments for terpenes and terpenoids, increasing yields further^{15,14}. Co-production of terpenes and lipid droplets was shown to not only increase terpene yields, but also enable a platform for bioproduction of both terpenes and other lipids of interest, such as triacylglycerols of which the lipid droplets are composed. These co-production strategies for terpenoids and lipid droplets may further improve economic feasibility of bioproduction hosts, which has become a focus with the green alga *Haematococcus pluvialis* producing the tetraterpenoid astaxanthin and triacylglycerols^{119,120}.

In this work, I have taken a multi-pronged approach to advance plants as a production platform for squalene and triterpenoid derivatives. A series of enzyme screenings were performed to optimize plastidial targeted squalene biosynthesis, using both native and engineered enzyme variants. The Lipid Droplet Surface Protein from *Nannochloropsis oceanica* (NoLDSP)¹²¹ was used to anchor the optimized squalene pathway to the surface of cytosolic lipid droplets in different variations, synthesizing squalene at the surface of lipid droplets and increasing yields. Next, the lipid droplet scaffolding strategy was re-targeted from the cytosol to plastids, where scaffolding occurred on membranes throughout chloroplasts and ameliorated negative effects of squalene accumulation on photosynthesis. Finally, combinations of *AtIPK*, *RcMPD*, and *AtBCCP1* were co-expressed with cytosolic and lipid droplet squalene pathways in attempts to boost yields further.

Results and Discussion

Screening to improve the entry step in the MEP pathway

The entry step in the MEP pathway synthesizes 1-deoxy-d-xylulose-5-phosphate (DXP) from pyruvate and GAP, catalyzed by DXS (Figure 2.1). With this step being a major limiting step in the MEP pathway¹¹⁴, controlling flux through the pathway³⁷, as well as being feedback inhibited by the end products IDP and DMADP¹¹⁸, it has become a target for increasing terpenoid production. Previous work has created novel, bacterial DXS-like enzymes (nDXSs) to synthesize DXP from an alternative substrate⁴⁰, or mutate DXS enzymes from poplar to de-regulate and reduce feedback inhibition³⁵. The nDXSs were shown to complement *dxs* knockout lines of *Escherichia coli* grown on xylose as the sole carbon source, synthesizing DXP from ribulose 5-phosphate (Ru5P) (Figure 2.1). Furthermore, it was shown that fusing the nDXSs to the next enzyme in the MEP pathway, DXS reductoisomerase (DXR), further increased flux through the pathway in *E. coli*⁴⁰. In studying DXS feedback inhibition, another study found a double mutant from a *Populus trichocarpa* DXS (*PtDXS* A147G:A352G) had reduced feedback inhibition by IDP and DMADP *in vitro*³⁵. In addition to reduced feedback inhibition, *PtDXS* A147G:A352G showed reduced activity *in vitro*, but this was never tested in planta to determine whether the reduced feedback inhibition can overcome the reduced activity when overexpressed. The two most successful nDXS enzymes from *E. coli*, RibB G108S and YajO, the *P. trichocarpa* double mutant *PtDXS* A147G:A352G, the wild type *PtDXS*, and a DXS from *Coleus forskohlii* (*CfDXS*) (Table A.1) were included in this study. In attempts to overcome limitations at this entry step, introduce novel contributions to the MEP pathway, and

increase terpene production in plants, the native and mutant DXSs, and bacterial nDXSs, alone and fused with DXR, were screened for terpene yields (Figure 2.2a).

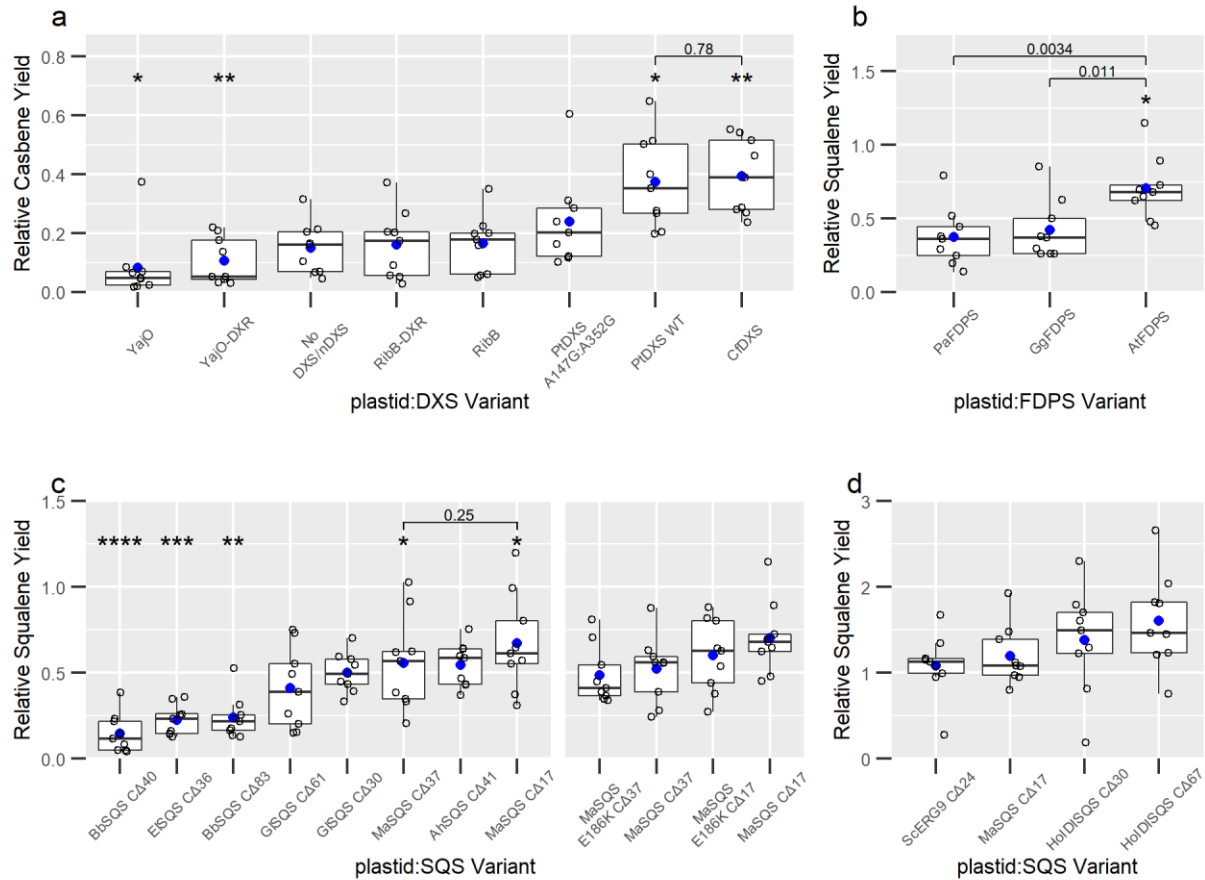


Figure 2.2: Plastid targeted pathway optimization for squalene production. Screening of key steps in plastid localization for DXS and nDXS (a), FDPS orthologues (b), truncated SQS orthologues (c, left), *MaSQS* mutants (c, right), and additional SQS variants (d). Each panel presents data collected from separate transient expression experiments, except panel (c, right) was from the same experiment as panel (b) while panel (c, left) was a separate experiment. Open circles are individual data points, blue circles are mean value, and horizontal line within box represents the median value. The box shows the range from the lower 25th percentile to the upper 75th percentile. The upper and lower whiskers extend to the largest and smallest data point no further than 1.5x the inter-quartile range, with points lying outside the whiskers considered outliers. Asterisks indicate significant difference of the mean relative squalene yield for each variant compared to mean of all variants in that experiment based on *t* test. ‘*’: $p \leq 0.05$; ‘**’: $p \leq 0.01$; ‘***’: $p \leq 0.001$; ‘****’: $p \leq 0.0001$. Individual statistical comparisons between means are shown by brackets and the indicated *p*-value.

While the plant DXS enzymes contain a native plastid transit peptide, a transit peptide sequence from the *Arabidopsis thaliana* Rubisco small subunit¹²² was added to the N-termini of the bacterial nDXSs, targeting these proteins to plastids. Using *Agrobacterium*-mediated, transient expression in *Nicotiana benthamiana*, each candidate gene was co-expressed along with a geranylgeranyl diphosphate synthase gene from *C. forskohlii* (*CfGGDPS*) and a casbene synthase gene from *Daphne genkwa* (*DgCasS*) (Table A.1) to synthesize the diterpene casbene as a proxy for flux towards a terpene product. While *CfDXS* and *PtDXS* demonstrated the highest casbene yields, *PtDXS* A147G:A352G showed reduced yields and the nDXSs did not show any increase over control lines of *CasS* and *GGDPS* only (Figure 2.2a).

CfDXS, the highest yielding DXS, increased casbene yields by 163% over control lines and showed 80.9% higher yields than the mean of all DXS and nDXS variations (Figure 2.2a). The wild type *PtDXS* showed 72% higher casbene yields than the mean of all variations and showed 57% higher yields than *PtDXS* A147G:A352G. While Banerjee et al.³⁵ demonstrated *PtDXS* A147G:A352G had reduced feedback inhibition, they also concluded the enzyme had reduced activity *in vitro*. The experiments here suggest the reduction in feedback inhibition does not overcome the reduced enzyme activity to increase terpene yields in plants. The nDXSs did not increase casbene yields, although the substrate, Ru5P, is typically present in chloroplasts as an intermediate of the Calvin-Benson cycle¹²³. While the nDXSs may provide an alternative route capable of complementing a DXS knockout in *E. coli*, nDXS gene overexpression did not result in more terpene accumulation than overexpression of the other wild type DXS variants in *N. benthamiana*. The wild type *CfDXS* resulted in the highest relative casbene yields and

was used in subsequent engineering and screening of FDPS and SQS candidates to improve plastidial squalene yields.

SQS and FDPS screening to improve squalene yields

While overexpression of key genes involved in the MVA or MEP pathways improves general terpenoid production, optimizing the downstream reactions towards specific products further increases yields^{41,124}. Here, screening of diverse orthologs and engineered variants of FDPS and SQS enabled selection of an optimal combination for squalene production. Six of the orthologous SQS genes were codon optimized for expression in *N. benthamiana* from the following organisms: *Amaranthus hybridus* (*AhSQS*), *Botryococcus braunii* (*BbSQS*), *Euphorbia lathyris* (*ElSQS*), *Ganoderma lucidum* (*GlSQS*), *Mortierella alpina* (*MaSQS*) and *Saccharomyces cerevisiae* (*ScERG9*), all species that can accumulate large amounts of squalene-related compounds. Additionally, a mutant of *MaSQS E186K* was created that was previously shown to improve catalytic efficiency 3.4-fold in *in vitro* studies¹²⁵. Finally, two truncated SQS variants from the diatom *Haslea ostrearia* (*HoDISQS*), which is a native fusion gene encoding an isopentenyl diphosphate isomerase (IDI) fused to the N-terminus of the SQS, were included. Both protein domains of *HoDISQS* were shown to be functional¹²⁶, and previous studies have shown co-expression of the gene encoding IDI, which catalyzes interconversion of IDP and DMADP, increases terpene yields^{126,127}. Screenings of SQS and FDPS variants were performed with plastid targeting, to compartmentalize squalene accumulation and avoid influence on native, cytosolic squalene biosynthesis.

To target SQS candidates to plastids, first the predicted C-terminal signal peptide which anchors the protein to the endoplasmic reticulum was removed to solubilize the

protein. Each candidate was truncated by two different lengths based on amino acid sequence alignment (truncation length indicated by Δ). The larger truncation was 10 amino acids following the end of conserved homology between sequences, and the shorter eliminated about half the number of amino acids. The same *A. thaliana* transit peptide was then added to the N-termini of truncated variants, targeting solubilized SQS candidates to plastids. SQS candidates were co-expressed with plastid targeted *FDPS* from *A. thaliana* (*AtFDPS*) and *CfDXS*, then squalene yields were measured, which are reported as a relative ratio of squalene to the added internal standard, n-hexacosane (Figure 2.2c). Expression of only plastid targeted *AtFDPS* and *CfDXS* resulted in squalene levels similar to background, indicating little influence on cytosolic squalene production (Figure A.1). Candidate *FDPS* genes were compared from three species: *A. thaliana*, *Picea abies*, and *Gallus gallus* (*N. benthamiana* codon optimized). Each *FDPS* gene was co-expressed with *CfDXS* and *MaSQS CΔ17* and squalene was measured (Figure 2.2b).

Comparing *FDPS* variants, *AtFDPS* had statistically significant higher yields than both *PaFDPS* and *GgFDPS* ($p < 0.01$ and $p < 0.05$, respectively), which was 41% higher than the mean of all *FDPS* variants (Figure 2.2b). Compared to the mean squalene yield of all variants, *BbSQS CΔ40*, *BbSQS CΔ83*, and *ElSQS CΔ36* had statistically significant lower yields ($p < 0.001$, $p < 0.05$, and $p < 0.05$, respectively), while *AhSQS CΔ41* and *MaSQS CΔ17* had significant higher yields ($p < 0.05$) (Figure 2.2c). The second *AhSQS* truncation, *AhSQS CΔ20*, was compared to *MaSQS CΔ17* and both resulted in similar yields (Figure A.1a). The *HoIDISQS* truncated variants were screened alongside *MaSQS CΔ17* and the commonly used yeast SQS, *ScERG9*, which showed a slight, but not

significant, increase in squalene yields (Figure 2.2d). Although these are not directly comparable without creating IDI fusions with other variants, these fusions may be worth further investigation in future engineering. While the *MaSQS* CΔ17 E186K variant previously demonstrated increased catalytic efficiency (k_{cat}/K_m) *in vitro*¹²⁵, it did not increase squalene yields with either truncation variant in this system (Figure 2.2c). *MaSQS* CΔ17 showed squalene yields 63% higher than the mean yield of other variants (Figure 2.2c) and was chosen as the SQS to use for development of lipid droplet scaffolding, along with *AtFDPS*.

Lipid droplet scaffolding optimization

Previous work demonstrated terpene synthases can be targeted to the surface of lipid droplets by fusing the enzymes to *NoLDSP*, where terpenoids are stored within the lipid droplets¹⁵. Unlike other commonly employed lipid droplet proteins like oleosins and seipins which can also be used for scaffolding^{128,129}, *NoLDSP* has no known plant orthologs¹²¹. Additionally, *NoLDSP* does not rescue oleosin functions of triacylglycerol turnover, possibly due to a lack of species-specific protein recruitment, which may be favorable in the context of lipid droplet overproduction¹²¹. Here, *AtFDPS* and *MaSQS* CΔ17 were used to re-localize squalene biosynthesis to the surface of cytosolic lipid droplets through fusions to *NoLDSP*. SQS and FDPS gene variants were co-expressed with the gene coding for a truncated form of HMGR, to reduce feedback inhibition, from *Euphorbia lathyris* (*ElHMGR*¹⁵⁹⁻⁵⁸²) that was previously shown to drive flux through the MVA pathway and increase cytosolic terpenoid yields¹⁵. To increase lipid droplet formation, the gene for a C-terminal, truncated WRINKLED1 transcription factor from *Arabidopsis thaliana* (*AtWRI1*¹⁻³⁹⁷) was co-expressed, which has been shown to activate

fatty acid biosynthetic pathways and oil production^{130,131}. Co-expression of the genes for *AtWRI1* and *NoLDSP*, with or without fusions to terpene synthases, induces accumulation of lipid droplets^{15,121}, which is utilized here to overproduce and functionalize lipid droplets as synthetic organelles. It was previously demonstrated that terpenes are effectively sequestered within lipid droplets when the biosynthetic pathways are anchored to the surface *NoLDSP* fusions¹⁵.

Several fusion variants were created to target *AtFDPS* and *MaSQS CΔ17*, together or separately, to lipid droplets (Figure 2.3a and Figure A.1c). Replacing the SQS endoplasmic reticulum retention signal, the N-terminus of *NoLDSP* was fused to the C-terminus of the truncated SQS (SQS-LDSP). Since FDPS enzymes are natively soluble, LDSP was initially fused to either the N- or C-terminus of *AtFDPS* (LDSP-FDPS or FDPS-LDSP, respectively). The FDPS-LDSP fusion demonstrated reduced yields (Figure A.1c), possibly because of interference of the C-terminal fusion located near the active site, according to the crystal structure of the human FDPS¹³². Therefore, all other combinations used variations of *NoLDSP* fused to the C-terminus of *MaSQS CΔ17* and the N-terminus of *AtFDPS*.

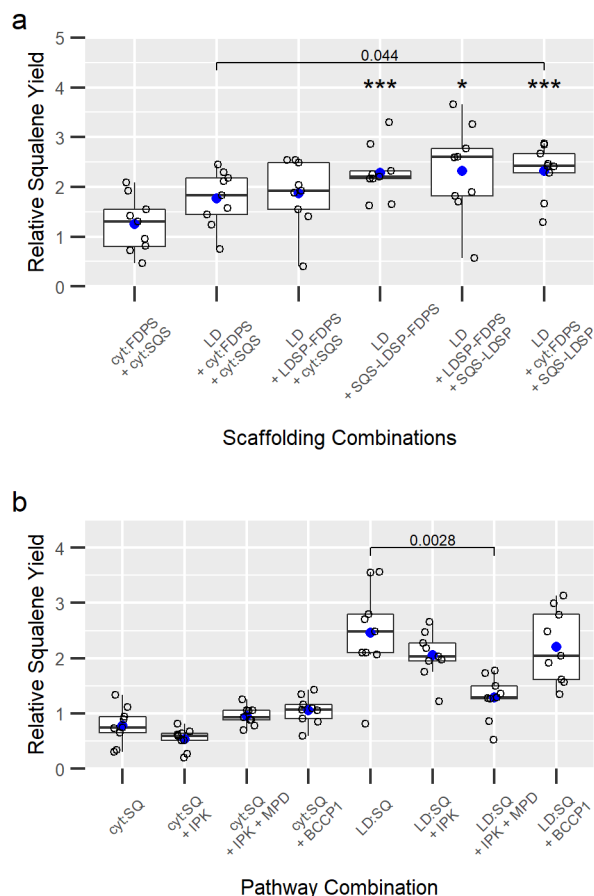


Figure 2.3: Screening of lipid droplet scaffolding combinations (a) and compatibility of alternative pathway contributions to engineered squalene pathways in the cytosol (b). All samples are co-expressed with *EIHMGR*¹⁵⁹⁻⁵⁸² and samples include “LD” if they co-express *AtWRI1*¹⁻³⁹⁷ and either *NoLDSP* alone or with the indicated *NoLDSP* fusion. Open circles are individual data points, blue circles are mean value, and horizontal line within box represents the median value. The box shows the range from the lower 25th percentile to the upper 75th percentile. The upper and lower whiskers extend to the largest and smallest data point no further than 1.5x the inter-quartile range, with points lying outside the whiskers considered outliers. Statistical significance compared to the mean of the cytosolic squalene pathway without lipid droplet scaffolding (cyt:FDPS + cyt:SQS) based on *t* test is indicated by an asterisk: “*”: *p* ≤ 0.05; “***”: *p* ≤ 0.001. Individual comparisons between variables are indicated by brackets with the corresponding *p*-value.

Each fusion protein, as well as soluble cytosolic versions of SQS and FDPS (cyt:SQS and cyt:FDPS), were used to test co-production of squalene and lipid droplets. The co-production of lipid droplets (*AtWRI1* + *NoLDSP*) with the soluble, cytosolic pathway *EIHMGR*¹⁵⁹⁻⁵⁸² + cyt:SQS + cyt:FDPS increased mean squalene yields by 41%,

with a similar increase of 49% when only FDPS was anchored to lipid droplets as LDSP-FDPS (Figure 2.3a). A significant increase in squalene yields of more than 80% was observed in lipid droplet scaffolding combinations involving SQS fusions to LDSP, which included cyt:SQS-LDSP + cyt:FDPS, cyt:SQS-LDSP + cyt:LDSP-FDPS, and cyt:SQS-LDSP-FDPS.

These data suggest it is key for the final step in the pathway, SQS, to be anchored to the surface of lipid droplets to increase yields. This may enable direct lipid droplet interaction for the squalene as it is synthesized. Co-localization of both FDPS and SQS at the surface of the droplets resulting in significant yield increases demonstrates a method to create synthetic organelles, which may be an effective strategy to direct biosynthesis further towards higher-value, squalene-derived triterpenoids, or other classes of products. Manipulation of lipid droplet architecture may provide an additional route to further modify scaffolding and production.

Targeting LDSP scaffolds to plastids

Further sub-compartmentalization of pathways within plastids may provide another strategy to re-direct accumulation of products like squalene¹⁴. The SQS-LDSP-FDPS fusion protein was targeted to plastids, through the addition of a transit peptide, to determine whether the pathway would remain functional if scaffolded in plastids. A plasmid was created where plast:*CfDXS*, plast:*AtFDPS*, and plast:*MaSQS CΔ17* are each separated by an LP4/2A linker^{133,104} (pDFS). The LP4/2A is a hybrid linker which combines post-translational cleavage of LP4 with the co-translational “cleavage” of 2A, allowing expression of a single transcript while producing separate protein products. When compared to pDFS, co-expression of plast:*CfDXS* with the fusion of plast:SQS-

LDSP-FDPS resulted in similar squalene yields (Figure A.1b). To determine if the *plast:SQS-LDSP-FDPS* fusion was successfully targeted to plastids, vectors were constructed to use EYFP and an EYFP-*NoLDSP* fusion, both with and without the plastid transit peptide. A series of experiments were then performed to determine which chloroplast associated membranes the *plast:EYFP-NoLDSP* fusion proteins were localizing to.

First, confocal microscopy was performed on *Agrobacterium*-infiltrated *N. benthamiana* containing constructs for either an empty vector, *cyt:EYFP*, *cyt:EYFP-NoLDSP*, *plast:EYFP*, or *plast:EYFP-NoLDSP*. It was previously reported and confirmed with Nile Red staining that *NoLDSP* overexpression induces lipid droplet formation¹⁵. This is confirmed here where the *cyt:EYFP-NoLDSP* fusion can be seen aggregating to cytosolic droplets (Figure A.2). In plants expressing the genes for *plast:EYFP-NoLDSP*, the fusion protein appears to aggregate along the plastid envelopes as well as forming a distinct punctate pattern within chloroplasts (Figure 2.4). This suggested *plast:EYFP-NoLDSP* localization to multiple membranes within chloroplasts, including plastoglobule-like structures. In the *plast:EYFP-NoLDSP* lines, likely cytosolic lipid droplet structures are also seen, suggesting *NoLDSP* may still be forming lipid droplets before the fusion protein can be transported to plastids. Future engineering of chloroplast genomes for plastidial expression of *LDSP* fusion genes may prevent cytosolic localization and induction of lipid droplet formation.

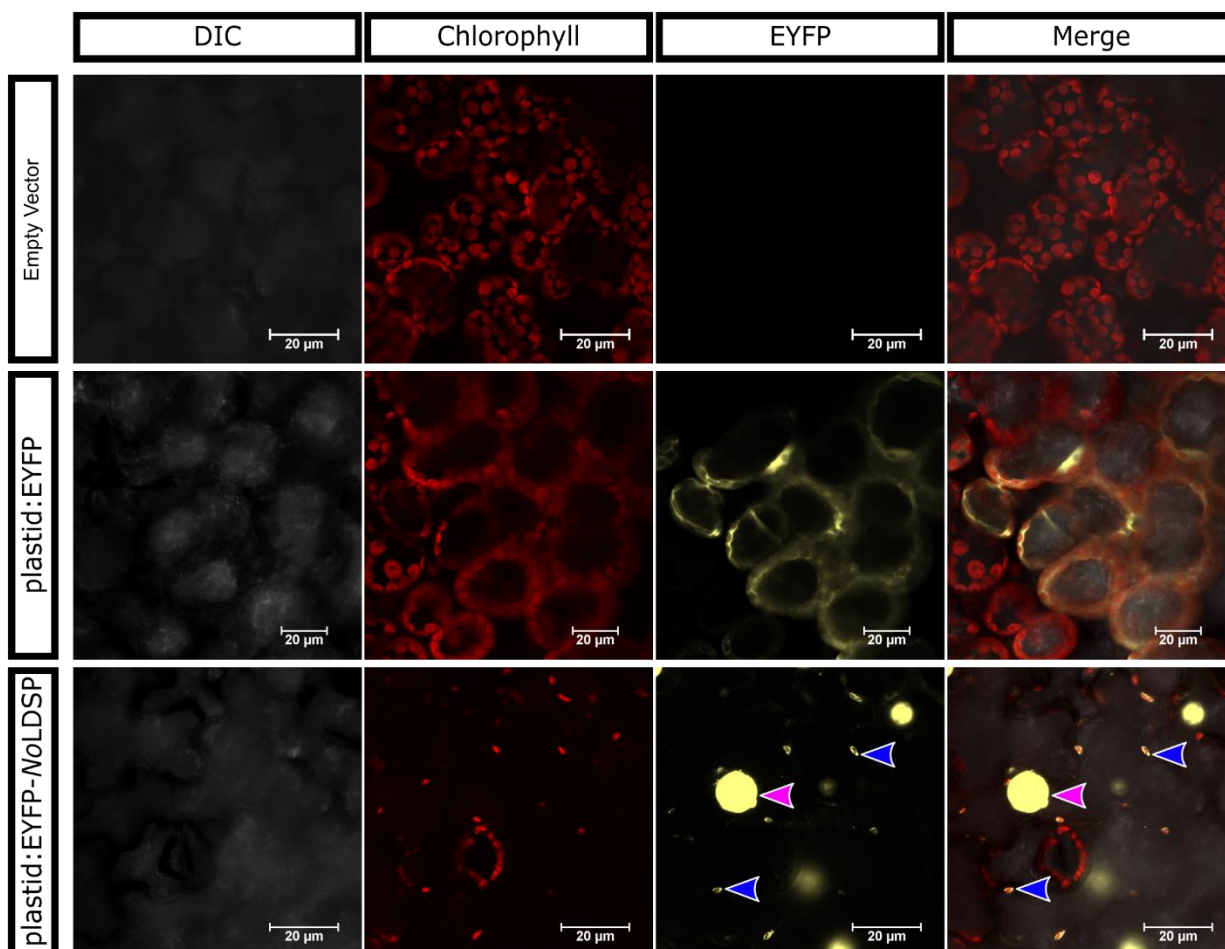


Figure 2.4: Confocal microscopy of plastid targeted EYFP and EYFP-NoLDSP to compare membrane localization. EYFP fluorescence and chlorophyll autofluorescence were measured with excitation:emission wavelengths of 513.9 nm:585 nm and 561 nm:700 nm, respectively. Blue arrows point to EYFP punctate seen in chloroplasts; pink arrows point to EYFP aggregating at cytosolic lipid droplets still seen being formed.

Previous studies have demonstrated the ability to isolate chloroplasts, separate the stroma, and fractionate plastidial membranes to study proteins associated with each membrane type^{134,135}. To determine localization of plast:EYFP-NoLDSP fusion proteins within chloroplasts, further analysis was performed here through chloroplast isolation and membrane fractionation with whole-plant, vacuum infiltrated lines expressing plast:*EYFP* or plast:*EYFP-NoLDSP*. Chloroplasts were isolated, lysed, and membranes separated from the stroma into three fractions (Figure 2.5a and Figure A.3a). The plast:EFYP-

NoLDSP fusion proteins were found associating with the stroma and all three membrane fractions, determined by anti-GFP antibody (Figure 2.5b and Figure A.3b). Also seen in each fraction is what may be a cleavage product of the plast: EYFP-NoLDSP fusion protein closer to the size of EYFP, though further analysis would be needed to confirm.

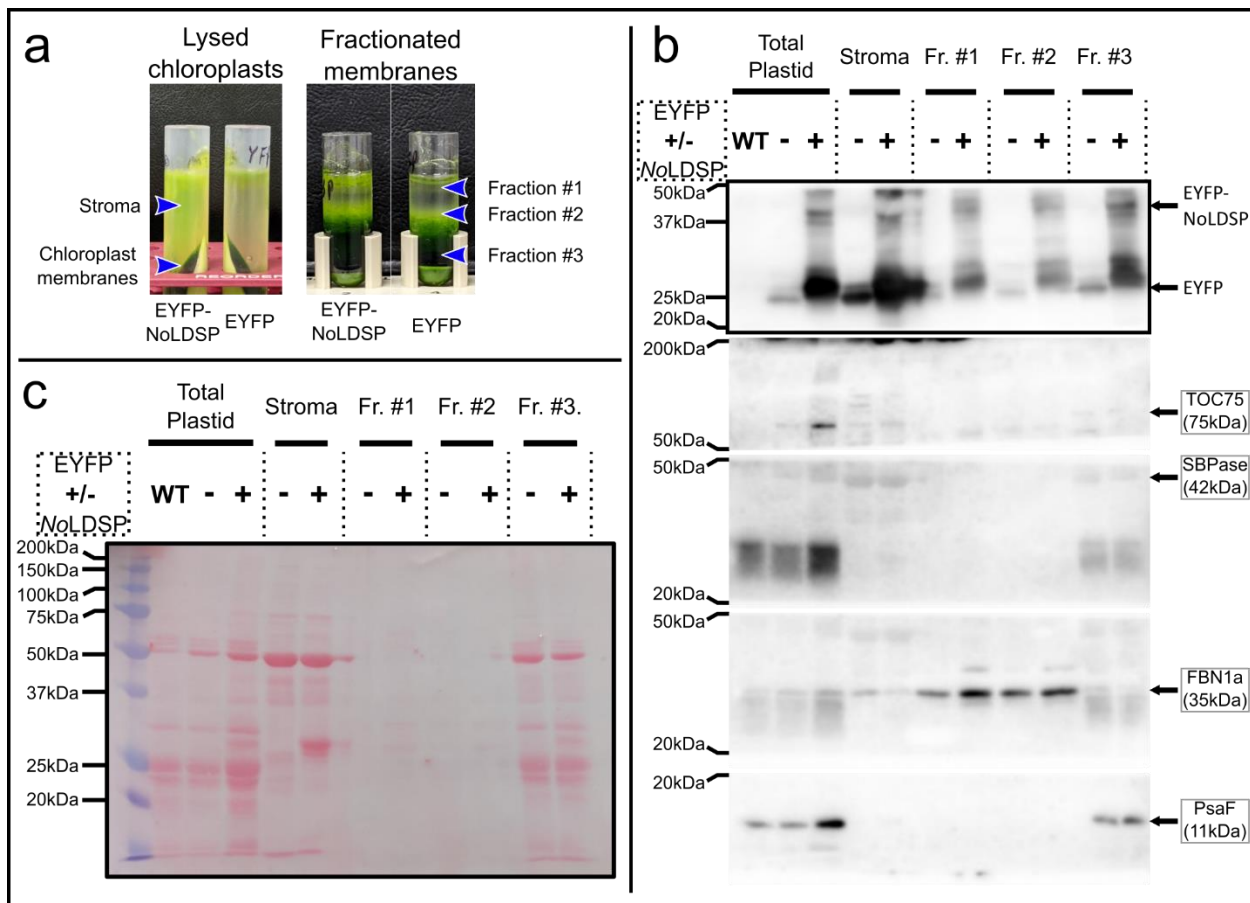


Figure 2.5: Western blots to determine membrane localization of plastid target, EYFP-NoLDSP fusion proteins. Chloroplast fractions (a) are labeled where the different samples were taken. Western blots using antibodies specific to proteins from each membrane fraction (b) are labeled with the expected protein mass and the ladder markers where the membrane was cut for antibody application. The in-tact membrane stained with Ponceau S dye (c) prior to cutting fragments for antibody visualization. Each lane is indicated as uninfiltrated, wild type plants (WT) or the presence of EYFP with (+) or without (-) fusion to NoLDSP. Arrows indicate bands at the expected sizes of each protein of interest. In (b), the membrane fragment cut at 50 kDa and 20 kDa was first visualized with anti-FBN1a followed by anti-SBPase. To analyze EYFP and EYFP-NoLDSP, the 50 kDa – 20 kDa membrane was then visualized using an anti-GFP antibody.

The stromal fraction was confirmed by application of the antibody for sedoheptulose-1,7-bisphosphatase (anti-SBPase) (Figure 2.5b and Figure A.3b). Fractions #1 and #3 were confirmed to contain plastoglobules and thylakoids, respectively, as determined by antibodies for fibrillin 1a (anti-FBN1a) and a subunit of photosystem I (anti-PsaF) (Figure 2.5b and Figure A.3b). Fraction #2 was inconsistent between experiments when visualized with antibodies for the 75 kDa subunit of the Translocon of the Outer Chloroplast (anti-TOC75), a protein associated with the outer membrane of the chloroplast envelope. TOC75 was not detected in Figure 2.5b but is detected in Figure A.3b, suggesting this fraction contains chloroplast envelopes. In support of these findings, a previous study visualized a similar fraction with anti-TOC75 and determined it to be predominantly chloroplast envelopes¹³⁴. Plastidial targeting of NoLDSP fusion proteins, therefore, enables non-specific, membrane scaffolding of proteins and pathways throughout the chloroplast.

Incorporating alternative contributions for the MVA pathway

To determine if squalene yields could be improved by introducing alternative contributions to the MVA pathway, further gene screenings were performed. While archaea rely on the MVA pathway, most lack the final two enzymes of the classical pathway to form IDP¹³⁶. In the classical MVA pathway, phosphomevalonate (MVAP) is phosphorylated by MVAP kinase (PMK) to form mevalonate diphosphate (MVADP), followed by decarboxylation by MVADP decarboxylase (MDD) to form IDP. In this alternative pathway (Figure 2.1), MVAP is first decarboxylated by phosphomevalonate decarboxylase (MPD) to form isopentenyl phosphate (IP), which is then phosphorylated to IDP with an IP kinase (IPK).

Recently, cytosolic pools of IP were found and IPK orthologs identified in plants³⁰, suggesting a role of IPK in regulation of IDP concentrations. While no MPD orthologs have been discovered in plants, co-expression of the MPD gene from the bacterium *R. castenholzii* (*RcMPD*) with *A. thaliana* IPK gene (*AtIPK*) showed increased production rates of terpenoids in the *Nicotiana tabacum* transient expression system³¹. This alternative pathway towards IDP was tested here to determine compatibility with the lipid droplet scaffolding strategy and to possibly increase squalene yields further.

In a separate approach to increase yields of cytosolic terpenoid pathways, overexpression of a biotin carboxyl carrier protein gene from *A. thaliana* (*AtBCCP1*) was investigated. It was previously shown that high accumulation of BCCP1 can disrupt the acetyl-CoA carboxylase complex, reducing conversion of acetyl-CoA to malonyl-CoA, the committed step towards plastidial *de novo* fatty acid biosynthesis³⁶. Disrupting the acetyl-CoA carboxylase complex in tandem with overexpression of terpenoid pathways led to increased yields of cytosolic terpenoids, presumably by allowing more acetyl-CoA to enter the MVA pathway (Figure 2.1).

These genes were co-expressed with the cytosolic squalene pathway both with and without lipid droplet scaffolding for comparison (Figure 2.3b). The addition *AtIPK* alone or *AtIPK* and *RcMPD* did not increase squalene yields. When combined with the lipid droplet scaffold, the addition of both *AtIPK* and *RcMPD* significantly reduced squalene yields. In this work, *AtIPK* and *RcMPD* were tested in an entirely transiently expressed system to synthesize the non-volatile squalene. The initial work was performed by transiently expressing the pathway for the volatile sesquiterpene santalene in a transgenic *RcMPD* overexpression line of *Nicotiana tabacum*³¹. The rates of santalene

emission were increased with overexpression of *RcMPD* and *AtIPK*, but overall yields were not measured. While the use of these enzymes may increase rates of terpenoid production, they did not increase overall squalene yields in this study. There was also no increase in squalene yields when *AtBCCP1* was added to transient expression systems employed. The initial work characterizing the role of *AtBCCP1* in cytosolic terpene production showed increased yields of the sesquiterpene bisabolol when overexpressing bisabolol synthase, *HMGR*, and *BCCP1*. While *AtBCCP1* overexpression, as well as *AtIPK* and *RcMPD*, did not increase squalene yields here, this may be due to limitations in storage capacity rather than enzyme activities. The increase in squalene yields when lipid droplets are co-produced (Figure 2.3b) demonstrates metabolic capability to produce more squalene, but may require greater storage capacity.

Investigating how expression of squalene pathways affect photosynthesis

Engineered biosynthetic pathways and accumulation of products in plants have been reported to hinder overall productivity and result in negative phenotypes such as stunted growth and reduced photosynthesis^{13,14,24,111,112}. To evaluate possible effects of these engineered pathways on native physiology, a series of gas-exchange experiments were performed under various squalene producing conditions. Although pathways were developed here using transient expression in vacuum infiltrated *N. benthamiana*, these experiments provide insight into how they may affect stably transformed crops. Four squalene strategies were evaluated for their effect on photosynthesis as an indicator for influences on plant physiology: (i) cyt:SQ⁽⁻⁾ LDSP (cyt:*E/HMGR* + cyt:*AtFDPS* + cyt:*MaSQS* CΔ17), (ii) cyt:SQ⁽⁺⁾ LDSP (*AtWRI1* + cyt:*E/HMGR* + cyt:*SQS-LDSP-FDPS*), (iii) plast:SQ⁽⁻⁾ LDSP (plast:*CfDXS* + plast:*AtFDPS* + plast:*MaSQS* CΔ17), and (iv) plast:SQ⁽⁺⁾ LDSP

(plast:CfDXS + plast:SQS-LDSP-FDPS). Each strategy reduced overall photosynthesis on both days 3 and 5, compared to pre-infiltration measurements (Figure 2.6a). Plants infiltrated with 200 μ M acetosyringone in water or with *Agrobacterium* containing an empty pEAQ-HT vector showed no significant differences in photosynthesis and squalene yields compared to un-infiltrated plants (Figure A.4). This indicates effects on leaf physiology reported here were due to expression of genes involved in the various squalene pathways studied.

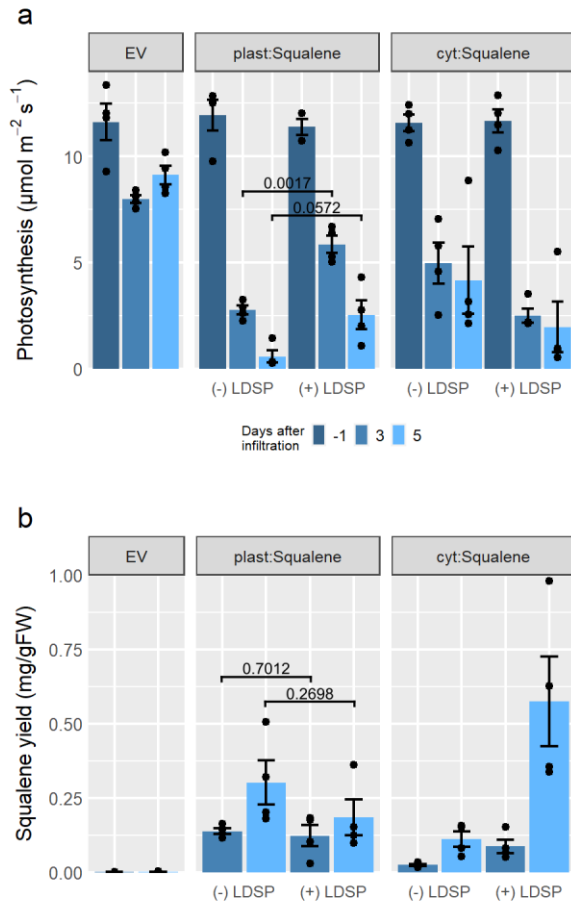


Figure 2.6: Comparison of photosynthesis and squalene biosynthesis in leaves transiently expressing cytosolic and plastid targeted squalene pathways. Photosynthesis data (a) and squalene yields (b) for each set of plants compared between the empty vector (EV), plastid squalene pathways (plast:Squalene) with (+LDSP) and without (-LDSP) membrane scaffolding, and cytosolic pathways (cyt:Squalene) with and without lipid droplet scaffolding. Black circles show individual data points and bars

Figure 2.6 (cont'd)

represent means $\pm$ standard error. $n = 4$ plants per treatment. Individual t test statistical comparisons between means are shown by brackets and the indicated p-value.

Both cytosolic strategies, with and without lipid droplet scaffolding, reduced photosynthesis on day 3, then further on day 5 (Figure 2.6a). Comparing the soluble pathways $\text{cyt:SQ}_{(-)}^{\text{LDSP}}$ and $\text{plast:SQ}_{(-)}^{\text{LDSP}}$, $\text{cyt:SQ}_{(-)}^{\text{LDSP}}$ pathway had 81% higher photosynthetic rates than $\text{plast:SQ}_{(-)}^{\text{LDSP}}$ on day 3 and 631% higher photosynthetic rates on day 5. Cytosolic lipid droplet scaffolding ($\text{cyt:SQ}_{(+)}^{\text{LDSP}}$) caused a greater reduction in photosynthesis than without, which is likely due to overexpression of *AtWRI1* as previous studies have shown this can induce severe downregulation of gene expression for proteins involved in the photosynthetic apparatus¹³⁰. Targeting the lipid droplet scaffold to plastids, however, moderated some of the negative effects on photosynthesis. Compared to the $\text{plast:SQ}_{(-)}^{\text{LDSP}}$ pathway on day 3, the $\text{plast:SQ}_{(+)}^{\text{LDSP}}$ infiltrated plants had 112% higher levels of photosynthesis and on day 5, $\text{plast:SQ}_{(+)}^{\text{LDSP}}$ had 345% higher levels of photosynthesis than $\text{plast:SQ}_{(-)}^{\text{LDSP}}$. These data demonstrate scaffolding of the plastidial, squalene biosynthetic pathway partially ameliorates the negative effects of squalene biosynthesis on photosynthesis.

The differences in photosynthesis seen in response to the expression of different squalene pathways was not due to variations in intercellular $[\text{CO}_2]$ resulting from differences in stomatal conductance (data not shown). We analyzed A/C_i curves (photosynthetic rate (A) plotted against intercellular CO_2 concentration, C_i) to understand what biochemical properties of photosynthesis are affected by squalene production (Figure A.5). Changes in rubisco activity (V_{cmax}) and ribulose 1,5-bisphosphate regeneration (or the rate of electron transport, J) followed similar trends and degrees of

change as seen for photosynthesis in leaves expressing different squalene pathways (Table A.2). This shows the negative effects on photosynthesis in leaves expressing squalene pathways were mainly due to reduced rubisco activity, and that targeting the LDSP scaffolds to plastids can moderate some of the negative effects on rubisco and photosynthesis. By comparing photosynthesis rates (Figure 2.6a) with squalene production (Figure 2.6b), it is clear that the increase in photosynthesis in plast:SQ₍₊₎ LDSP was due to scaffolding of the plastidial squalene biosynthetic pathway, and not due to a decrease in squalene production. The reason for how squalene negatively affects rubisco, and how scaffolding of the plastidial squalene biosynthetic pathway helps alleviate the negative effects on rubisco and photosynthesis can only be speculated at this time. It may be that squalene accumulation in plastids has a direct inhibitory effect on rubisco. Studies have demonstrated that squalene accumulation can form aggregates and interfere with native membranes^{137,138}. Compared to the soluble plastid pathway, the LDSP scaffolding in plastids may distribute the accumulation of squalene more broadly throughout various membranes and reduce disruption of protein organization on thylakoid membranes.

While squalene biosynthesis was detected in leaves expressing all squalene pathways tested, the largest increase was seen with the cytosolic strategy with lipid droplet scaffolding (Figure 2.6b). In comparison, squalene production by the plastidial pathways was less than half that of the cytosolic pathway with lipid droplet scaffolding. Additionally, when determining the amount of fixed carbon utilized in the engineered squalene pathways, plast:SQ₍₋₎ LDSP presented the highest conversion of 1.97% of fixed carbon towards squalene (Table 1), while all other strategies utilized less than 1% of fixed carbon. Compared to the microalgae *Botryococcus braunii*, in which upwards of 45% of

photosynthetic carbon is naturally directed towards terpenoid biosynthesis¹³⁹, there may be significant capacity to increase carbon partitioning towards squalene in these engineered systems.

Table 2.1: Summary of changes in squalene yields and photosynthesis when scaffolding squalene biosynthesis using NoLDSP. Changes are summarized for the plastid (plast:SQ) or cytosol (cyt:SQ) targeted pathways. Squalene production and photosynthetic carbon fixed during a 12 hour photoperiod were calculated in terms of μmol of carbon fixed per unit surface area (m^{-2}) of the leaf sampled, and the ratio was used to determine the percentage of photosynthetic fixed carbon converted to squalene (% of fixed carbon to squalene). Values represent means obtained from $n = 4$ plants per treatment.

Treatment	Day after infiltration	Squalene yield (mg/gFW)	Difference in squalene yield with LDSP	Difference in photosynthesis with LDSP	Squalene production ($\mu\text{mol}/\text{m}^2$)	Photosynthetic carbon fixed ($\mu\text{mol}/\text{m}^2$)	% of fixed carbon to squalene
Empty Vector	Day 3	9.49×10^{-4}	---	---	1.30	3.45×10^5	0.00%
Empty Vector	Day 5	1.66×10^{-3}	---	---	2.69	3.94×10^5	0.00%
plast:SQ (-) LDSP	Day 3	1.38×10^{-1}	- 11.2%	+ 112%	1.82×10^2	1.19×10^5	0.15%
plast:SQ (+) LDSP		1.23×10^{-1}			1.78×10^2	2.52×10^5	0.07%
plast:SQ (-) LDSP	Day 5	3.02×10^{-1}	- 38.9%	+ 345%	4.84×10^2	2.46×10^4	1.97%
plast:SQ (+) LDSP		1.84×10^{-1}			3.14×10^2	1.10×10^5	0.29%
cyt:SQ (-) LDSP	Day 3	6.28×10^{-2}	+ 67.6 %	- 49.8%	8.33×10^1	2.15×10^5	0.04%
cyt:SQ (+) LDSP		1.05×10^{-1}			1.44×10^2	1.08×10^5	0.13%
cyt:SQ (-) LDSP	Day 5	1.11×10^{-1}	+ 318 %	- 52.9%	1.48×10^2	1.80×10^5	0.08%
cyt:SQ (+) LDSP		5.75×10^{-1}			7.76×10^2	8.47×10^4	0.92%

In conclusion, optimizing key steps in squalene biosynthesis and compartmentalizing the pathway to cytosolic lipid droplets or within plastids and plastid membranes effectively improves production of squalene within plant systems. These strategies provide a platform which can be expanded to produce compounds directly formed from squalene like ambrein, or other squalene derived triterpenoids, sterols, and related bioproducts with important industrial applications. Combining these pathways with alternative precursor contributors without seeing increased yields suggests there may be need to further increase storage capacity. Additionally, there appears to be significant photosynthetic capacity to direct more fixed carbon to products. Both may be improved by further engineering of lipid droplet architecture or membrane scaffolding to increase overall storage capacity. Experiments here were performed using *Agrobacterium*-

mediated transient expression in *N. benthamiana*, which may inform future work to generate stable transformants.

Methods

Genes synthesized and cloned

Genes employed in this study are listed in Table A.1. Genes were either cloned from cDNA from the native host or synthesized as gene fragments from Integrated DNA Technologies (IDT) or Twist Bioscience. The IDT Codon Optimization Tool was used for genes codon optimized for expression in *N. benthamiana*. Genes were initially inserted into the pJET 1.2/blunt vector using the CloneJET PCR Cloning Kit (ThermoFisher Scientific). For transient expression experiments, genes were amplified with Phusion High-Fidelity DNA Polymerase (New England Biolabs) then inserted into the pEAQ-*HT* vector^{140,72} (digested with XhoI and NruI restriction enzymes) using the In-Fusion HD Cloning Kit (Takara Bio). The pEAQ-*HT* vector utilizes an enhanced expression platform using the Cauliflower Mosaic Virus 35S promoter, 5' and 3' untranslated regions from the Cowpea Mosaic Virus, and the nopaline synthase terminator. Additionally, this vector co-expresses the RNA silencing suppressor P19 gene using the 35S promoter and terminator from the Cauliflower Mosaic Virus. All genes following pJET 1.2/blunt and pEAQ-*HT* cloning were confirmed via Sanger sequencing provided by Psomagen, Inc. *Agrobacterium tumefaciens* LBA4404 cells were transformed with pEAQ-*HT* plasmids containing the gene of interest via electroporation then plated on LB media with 50 µg/mL of kanamycin and 25 µg/mL of rifampicin for selection. Transformed *Agrobacterium* cells were cultured overnight, and flash frozen in 20% glycerol for storage until needed for transient expression.

Agrobacterium-mediated transient expression, compound extraction, and measurement

Agrobacterium-mediated, transient expression experiments were performed similar to previously described methods¹⁵. Transformed *Agrobacterium* cells from 20% glycerol stocks were cultured in 5ml LB with 50 µg/mL kanamycin and 25 µg/mL rifampicin for 20 hr at 28 °C. These starter cultures were used to inoculate 25 mL of LB with 50 µg/mL kanamycin, also cultured for 20 hr at 28°C. Cultures were then centrifuged at 4,000 x g for 10 min, decanted, then re-suspended in 10 mL of water. This was repeated two more times for a total of three washes and finally re-suspended in 10 mL of water with 200 µM of acetosyringone, which was diluted to an OD₆₀₀ of 0.8 or 1.0 with 200 µM acetosyringone in water. Cultures for each set of DXS experiments were diluted to OD₆₀₀ of 0.8, while cultures for each set of squalene pathway optimization experiments were diluted to OD₆₀₀ of 1.0. Following dilution, cultures were shaken at 28 °C for 1-2 hr before infiltration. Cultures for co-infiltration were mixed in equal proportions and then syringe infiltrated into 3 leaves of 3 independent *N. benthamiana* plants (4-5 weeks old), for a total of 9 replicates of each combination in each experiment. Infiltrated plants were grown at 23-25°C with a 12-h photoperiod at 150 µmol m⁻² s⁻¹ for 5 days before extraction of compounds.

For casbene extractions, two 15 mm leaf discs were cut out and placed in a vial with hexane containing 20ng/µL ledol. Samples were left shaking overnight at room temperature, centrifuged at 525 x g for 20 min to pellet plant debris, and supernatant transferred to fresh amber GC vial for analysis. For squalene extractions, two 15 mm leaf discs were cut out and placed in 2 mL screw cap vials containing 0.1 mm glass beads

and one 3 mm tungsten carbide bead then flash frozen in liquid nitrogen and stored in -80 °C until extraction. Frozen leaf tissue samples were ground using a Qiagen TissueLyser at 30 rotations s⁻¹ for 1.5 min twice, 600 µL of hexane containing 50 ng/µL n-hexacosane added and samples vortexed, then samples were shaken for 2 hr at room temperature. 300 µL of water was added to aid with separation, samples centrifuged at 16,000 x g for 5 min, and organic layer transferred to amber GC vials for analysis.

Hexane extracts were analyzed via gas chromatography – flame ionization detection (GC-FID) on an Agilent 7890A and compared to retention times of a squalene standard. Peak areas for the internal standard, hexacosane, and squalene were extracted for comparison. To determine relative yields of squalene, or casbene, peak areas of squalene were divided by hexacosane, or ledol, peak areas, presenting a squalene or casbene yield relative to the internal standard. For quantification of squalene, fresh leaf tissue was weighed prior to extraction and a squalene, GC-FID calibration curve was created to determine yields. Squalene peak areas were normalized to the mean of hexacosane areas across samples and squalene quantified based on the calibration curve.

Plastid fractionation and western blots

Plastids from *Agrobacterium*-infiltrated leaves were extracted and fractionated similar to previously described methods¹³⁵. Whole *N. benthamiana* plants were vacuum infiltrated with *Agrobacterium* harboring pEAQ-HT vectors which contain either plast:EYFP gene or a plast:EYFP-NoLDSP fusion gene. Leaves from 15 full plants from each condition were blended in isolation buffer (330 mM sorbitol, 20 mM MOPS pH 7.6, 13 mM Tris-HCl, 3 mM MgCl₂, 0.1 % (w/v) BSA, 5 mM ascorbic acid, 5 mM reduced

cysteine, and a protease inhibitors mixture containing 74 μ M antipain, 130 μ M bestatin, 16.5 μ M chymostatin, 56 μ M E64, 2.3 μ M leupeptin, 37 μ M phosphoramidon, 209 μ M AEBSF, 0.5 μ M aprotinin, 50 mM NaF, 25 mM β -glycerophosphate, 1 mM Na-orthovanadate, and 10 mM Na-pyrophosphate) then filtered through Miracloth. Samples were centrifuged for 5 min at 1,500 x g at 4°C to pellet chloroplasts, then pellets were washed twice with washing buffer (50 mM HEPES pH 7.6, 5 mM ascorbic acid, 5 mM reduced cysteine, 330 mM sorbitol, and protease inhibitors mixture) and centrifuged again as described above. To lyse chloroplasts, washed samples were re-suspended and incubated for 30 min on ice in an osmotic shock buffer (10 mM Tricine pH 7.9, 1 mM EDTA, 0.6 M sucrose, and protease inhibitors mixture). Lysed chloroplasts were centrifuged for 1 hr at 100,000 x g to separate chloroplast stroma from membranes. Membrane pellet was resuspended in 2 mL of 48% sucrose in HE buffer (50 mM HEPES pH 7.9 and 2 mM EDTA), 1 mL transferred to an ultra-centrifuge tube where 800 μ L 5% sucrose in HE buffer was carefully overlaid, and the gradient centrifuged for 2 hr at 100,000 x g. The top, yellow layer consisting of plastoglobules, the second, yellow layer consisting of plastid envelopes, and the bottom, green layer consisting of thylakoids were removed for analysis (Figures 5a and Figure A.3a).

For each membrane layer, total protein was quantified using the Pierce™ BCA Protein Assay Kit (Thermo Scientific) and 5 μ g of total protein added to Laemmli sample buffer before boiling for 10 min. Each boiled sample was run on a 12% SDS gel with 4% SDS stacking gel to separate proteins, then proteins were transferred to a nitrocellulose membrane (Amersham™ Protran®), and total proteins were stained and visualized with incubation in Ponceau S dye (Figures 2.5c and Figure A.3c). The dye was removed and

the membrane incubated in blocking buffer (5% condensed milk in tris-buffered saline (TBS)) at room temperature for 1 hr. Membranes were washed in TBS with 0.1% Tween® 20 (TBS-T), cut at the specified molecular weight markers, then incubated at 4°C overnight with the appropriate antibody. The antibodies anti-SBPase, anti-TOC75, anti-PsaF, and anti-FBN1a were used to identify fractions for the stroma, plastid envelopes, thylakoids, and plastoglobules, respectively. Membranes were washed with TBS-T, incubated with the secondary, polyclonal anti-rabbit HRP antibody for 2 hr at room temperature, washed again then imaged using enhanced chemiluminescence. For membranes with EYFP, the secondary antibody was stripped, washed, and anti-GFP antibody applied and imaged as above, starting with blocking buffer.

Gas-exchange measurements

Nicotiana benthamiana plants for gas-exchange measurements were grown from seeds in Suremix (Michigan Grower Products) in 5" pots. Plants were grown under a 12-h photoperiod, a light intensity of 400 $\mu\text{mol m}^{-2} \text{s}^{-1}$, day/night temperatures of 25°C/20°C, and 60% humidity. Plants were kept in a growth chamber (Big-Foot, BioChambers) and fertilized using 1/2-strength Hoagland's solution¹⁴¹. 5-week-old plants were infiltrated with *Agrobacterium* harboring vectors for the indicated squalene pathways and controls at an OD₆₀₀ of 0.8 as described above, except using vacuum instead of syringe infiltration. For vacuum infiltration, whole plants were submerged in the respective *Agrobacterium* cultures, placed under a vacuum for 3-4 min, then vacuum quickly released to infiltrate leaves. The vacuum and release was repeated once to ensure full infiltration of leaves.

Gas exchange measurements described below were performed the day before infiltration of leaves, and on the 3rd and 5th day after infiltration. For each plant, gas

exchange measurements were performed on the 3rd fully expanded leaf counting from the top of the canopy. This leaf was tagged and the same leaf was measured consecutively prior to infiltration, and on the 3rd and 5th days after infiltration. For squalene measurements, on day 3, four 15 mm leaf discs were collected from the leaf immediately above the leaf being used for gas-exchange measurements. On day 5, leaf discs were collected from the leaf being measured following measurements at the end of the day. Squalene was extracted and measured as described above.

Photosynthetic rates (A), the operational efficiency of photosystem II in light adapted leaves (Φ_{PSII}), and stomatal conductance (g_{sw}) were measured simultaneously with the aid of a LI-6800 portable gas exchange system (LI-COR Biosciences, Lincoln, NE) connected to a Multiphase FlashTM Fluorometer (6800-01A). The environmental conditions inside the LI-6800 leaf chamber were set to match daytime growth chamber conditions: a light intensity of 400 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (50% blue light and 50% red light), temperature of 25°C, $[\text{CO}_2]$ of 400 $\mu\text{mol mol}^{-1}$ and water vapor content of 22 mmol mol^{-1} . First, the leaf was inserted into the leaf chamber and allowed to equilibrate for 30 min under the above conditions. A measurement was logged after photosynthesis reached steady state at the end of this equilibration period. Next, the light intensity inside the leaf chamber was then increased to 1000 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and the leaf was held under this saturating light condition until photosynthesis reached steady state. The response of photosynthesis to CO_2 was determined by measuring photosynthetic rates at varying $[\text{CO}_2]$. $[\text{CO}_2]$ was set to change from low to high (50, 100, 150, 200, 300, 350, 400, 450, 500, 550, 600, 700, 800, 1000, 1300, 1500 $\mu\text{mol mol}^{-1}$), and the leaf was allowed to equilibrate for 2-3 min at each CO_2 concentration before a measurement was logged.

These data were used to generate A/C_i curves. To determine the biochemical capacities underlying photosynthesis: maximum carboxylation rate (V_{cmax}), maximum rate of electron transport (J), triose phosphate utilization rate (TPU), A/C_i curves were fitted by the Farquhar-von Caemmerer-Berry biochemical model of photosynthesis^{142,143} using the following software: A/C_i curve fitting utility version 2.9 for tobacco^{143–145}.

Author contributions and acknowledgements

JDB and BRH conceived the study. JDB wrote the manuscript, with contributions by SMW for the gas-exchange measurement methods and results. DXS and nDXS screenings were performed by JDB and KR. Lipid droplet scaffolding screenings were performed by JDB and AB. FDPS, SQS, IPK, MPD, and BCCP1 screening, and plastid targeted scaffolding and fluorescence microscopy experiments were performed by JDB. Photosynthesis experiments were conceived and designed by JDB, BRH, TDS, and SMW and performed by JDB and SMW. Chloroplast fractionation experiments were conceived and designed by JDB, BRH, PKL, and REC and performed by JDB with close guidance from REC.

We would like to thank Balindile B. Motsa, Peiyen Kuo, Malik Sankofa, and Jade Lim for planting and maintenance of *Nicotiana benthamiana* plants used throughout this study, Dr. Noelia Pastor-Cantizano for assistance with fluorescence microscopy, and James Klug and Cody Keilen with the Michigan State University Growth Chamber Facility for maintaining facilities in which the plants were grown. This material is based upon work supported in part by the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Numbers DE-SC0018409 and DE-FC02-07ER64494. We would also like to acknowledge

partial support from the Department of Biochemistry and Molecular Biology startup funding and support from AgBioResearch (MICL02454). TDS received partial salary support from AgBioResearch. We collectively acknowledge that Michigan State University occupies the ancestral, traditional, and contemporary Lands of the Anishinaabeg – Three Fires Confederacy of Ojibwe, Odawa, and Potawatomi peoples. In particular, the University resides on Land ceded in the 1819 Treaty of Saginaw. We recognize, support, and advocate for the sovereignty of Michigan's twelve federally-recognized Indian nations, for historic Indigenous communities in Michigan, for Indigenous individuals and communities who live here now, and for those who were forcibly removed from their Homelands. By offering this Land Acknowledgement, we affirm Indigenous sovereignty and will work to hold Michigan State University more accountable to the needs of American Indian and Indigenous peoples. Figure 2.1 was created using Biorender.com.

CHAPTER 3

Engineered poplar for bioproduction of the triterpene squalene

Results from this chapter are being submitted for publication as part of the following manuscript:

Bibik, J. D., Kim, B., Sahu, A., Unda, F., Mansfield, S. D., Maravelias, C. T., Sharkey, T. D., Hamberger, B. R. Engineered poplar for bioproduction of the triterpene squalene.

Abstract

Building sustainable platforms for production of biofuels and specialty chemicals has become an increasingly important strategy to supplement and replace fossil fuels and petrochemicals. Terpenoids are the most diverse class of natural products which have many commercial roles as specialty chemicals. Poplar is a rapidly growing, biomass dense bioenergy crop with many species known to produce large amounts of the hemiterpene isoprene, suggesting an inherent capacity to produce large amounts of other terpenes. Here I aimed to engineer poplar with optimized pathways to produce squalene, a triterpene commonly used in cosmetic oils, a potential biofuel candidate, and the precursor to the diverse classes of triterpenoids and sterols. Squalene production was either re-targeted from the cytosol to plastids or co-produced with lipid droplets in the cytosol. Squalene and lipid droplet co-production appeared to be toxic, which I hypothesize to be due to disruption of adventitious root formation, suggesting a need for tissue specific production. Plastidial squalene production enabled up to 0.63 mg/g fresh weight in leaf tissue, which also resulted in reductions in isoprene emission and photosynthesis. These results were also studied through a technoeconomic analysis, providing further insight into developing poplar as a production host.

Introduction

Engineering plants for sustainable bioproduction of high-value chemicals has become of great interest⁵. One class of compounds of particular interest are terpenes and the further functionalized terpenoids, the most chemically diverse class of natural products. Terpenoid diversity throughout species has established their significance as major products in many herbal and medicinal plants used by humans for thousands of years¹¹. Many plants are known for naturally producing large amounts of terpenes and terpenoids, including mono- and sesqui- terpenoids in the Lamiaceae (mint) family, diterpenoid oleoresins in conifers, and isoprene in poplars¹⁴⁶. While these plants can produce these chemicals in relatively large amounts, scaling up for industrial production is often limited by a lack of plant biomass, accumulation of structurally similar compounds, low economic value, and many other factors. Furthermore, terpenoid diversity and complexity often makes them expensive and difficult, if not impossible, to chemically synthesize. With increased characterization of terpenoid biosynthesis in nature, biosynthetic pathways can be engineered into fast growing, non-food crop species as an inexpensive, ecologically sustainable, and larger scale alternative to chemical synthesis.

Terpenoids are derived from two common building blocks, dimethylallyl diphosphate (DMADP) and the isomer isopentenyl diphosphate (IDP), which are synthesized within plastids via the methylerythritol 4-phosphate (MEP) pathway or cytosolically from the mevalonate (MVA) pathway. The C₅ hemiterpenes, notably isoprene, are synthesized in plastids directly from one DMADP. The C₁₀ monoterpenes and C₂₀ diterpenes are also synthesized in plastids, using one DMADP and either one or three IDP molecules, respectively. Additionally, the C₄₀ tetraterpenes (including

carotenoids) are synthesized in plastids from two DMADP and six IDP molecules. Co-localized to the cytosol, or endoplasmic reticulum with access to cytosolic substrates, are the enzymes responsible for synthesis of the C₁₅ sesquiterpenes and the C₃₀ triterpenes. Many studies have developed strategies to engineer these pathways for increased production, re-direct biosynthesis between the cytosol and plastids, and even engineer co-production with or scaffolding on lipid droplets^{12–15}. Much of this research, however, has been performed using *Agrobacterium*-mediated transient expression. While recent efforts have made transient expression capable of producing terpenoids in gram-scale quantities²⁷, there remains great interest in developing transgenic crops for economically and environmentally sustainable bioproduction. It is therefore important to translate these strategies from transient expression to stably transformed, transgenic lines in species capable of sustainable bioproduction.

Due to its rapid growth, high lignocellulosic biomass, and established production for the pulp and paper industry, many poplar (*Populus* spp.) hybrids have become target feedstocks for production of biofuels and bioproducts⁶⁴. Bioengineering of poplar has been predominantly directed towards manipulation of the lignocellulosic content to improve conversion for pulp and paper, monolignol derived chemicals, or fermentable sugars^{147–149}. These fermentable sugars are then supplied as carbon sources for microbes, which in turn synthesize target biofuels and bioproducts, as opposed to the poplar directly producing the compounds. Recent analyses of another bioenergy feedstock crop, sorghum, have demonstrated potential to improve the economics of such systems by engineering the crops to directly produce higher-value chemicals, such as terpenes, which can be extracted prior to feedstock conversion⁶³. These platforms would

allow a more sustainable strategy for co-production of high-value chemicals and microbial feedstocks.

While most engineering of poplar has focused on production and conversion to monomeric sugars, there have been some examples of engineering for direct chemical production^{66,67}. One study engineered production of the phenylalanine-derived 2-phenylethanol, and its phenylethanol glucoside form, which accumulated in leaves and stems of a poplar hybrid⁶⁶. In an additional study, the same poplar hybrid was engineered with the pathways for either of the coniferyl alcohol derivatives eugenol or isoeugenol, which were further studied in 4-year field trials⁶⁷. These biosynthetic pathways were targeted because they are derived from the same precursors of the monolignols that form lignocellulosic biomass in poplar. In addition to monolignol production, poplars are major contributors to global emissions of the hemiterpene isoprene, demonstrating significant metabolic capacity for production of terpenoids, particularly in leaves. Therefore, with poplars demonstrating robust growth and biomass production, ability to naturally produce large amounts of terpenes (isoprene), and having well-established deployment for industrial use, it is a promising platform for sustainable terpenoid production.

In Chapter 2, I developed strategies to produce the C₃₀ triterpene squalene through pathway optimization and compartmentalization using transient expression in *Nicotiana benthamiana*. Squalene is commonly used in cosmetic oils, vaccines, and is a candidate biofuel in addition to being the precursor to higher-value triterpenoids with broad biotechnological applications. Squalene biosynthesis in plants natively occurs through condensation of one DMADP and two IDP molecules to form farnesyl diphosphate (FDP) by the soluble FDP synthase (FDPS), followed by further condensation of two FDP

molecules by the endoplasmic reticulum bound squalene synthase (SQS). The squalene production strategies were previously developed through transient expression to successfully re-target FDPS and SQS to plastids or the surface of lipid droplets (Figure 3.1).

I previously developed engineered squalene pathways to either re-target squalene biosynthesis from the cytosol to plastids or co-produce cytosolic lipid droplets with and without scaffolding of squalene biosynthetic enzymes at the surface (in Chapter 2). The first strategy used here in poplar engineering re-targets an *Arabidopsis thaliana* FDPS (AtFDPS) and a SQS from the fungal species *Mortierella alpina* with a 17 amino acid C-terminal truncation to remove the endoplasmic reticulum retention sequence (MaSQS CΔ17), to plastids by fusion with an N-terminal transit peptide from the *A. thaliana* Rubisco small subunit¹²² (Figure 3.1, right). To increase the amount of available IDP/DMADP in plastids, overexpression of the gene for 1-deoxy-d-xylulose-5-phosphate synthase from *Coleus forskohlii* (CfDXS) was included. As the entry step to the MEP pathway, DXS has been shown to be rate limiting and overexpression of the gene can overcome some of these limitations.

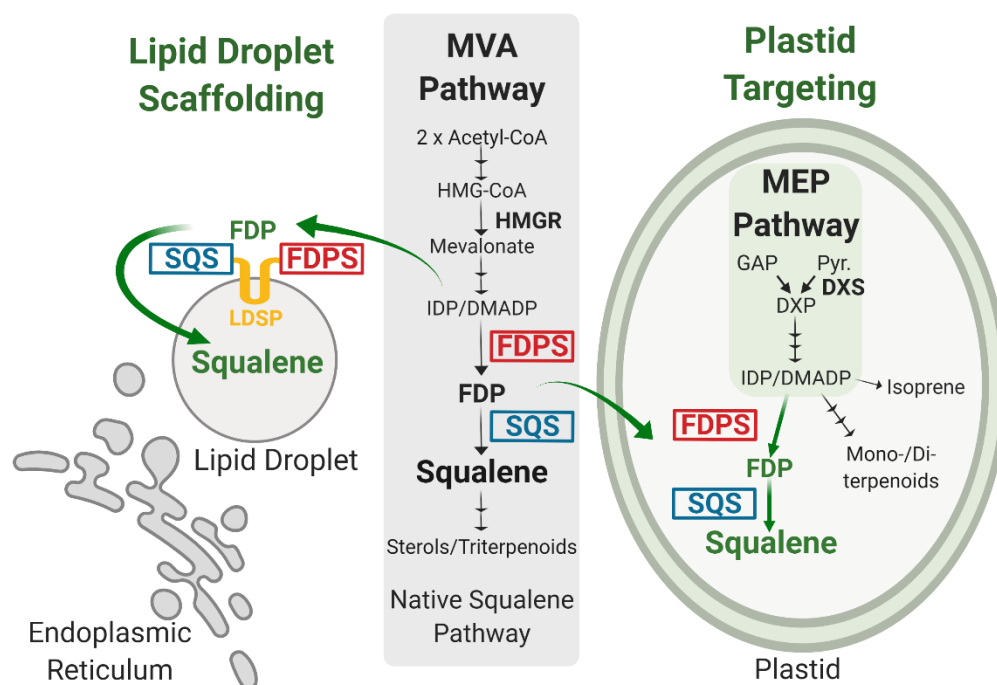


Figure 3.1: Representation of engineered squalene pathways used for poplar transformations. The enzymes required for biosynthesis of squalene, FDPS and SQS, were re-targeted to plastids (right) or used in combination with lipid droplet co-production or scaffolding through fusions with LDSP (left). Variations of both strategies were attempted when engineering poplar.

The second strategy tested here was designed for co-production of cytosolic lipid droplets and squalene also utilizing *At*FDPS and *Ma*SQS CΔ17 (Figure 3.1, left). To upregulate production of IDP/DMADP in the cytosol, the gene for the committed step to the MVA pathway, 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR), from *Euphorbia lathyris* (*El*HMGR¹⁵⁹⁻⁵⁸²) was included. Overexpression of the gene in truncated form of HMGR improves flux through the MVA pathway and increases cytosolic terpenoid yields¹⁵. In addition to the squalene biosynthetic pathway, this strategy co-produces lipid droplets through expression of the gene for truncated WRINKLED1 transcription factor from *A. thaliana* (*At*WRI1¹⁻³⁹⁷), which upregulates expression of pathways involved with fatty acid and lipid production^{130,131}, as well as the gene for Lipid Droplet Surface Protein

from *Nannochloropsis oceanica* (NoLDSP), which inserts into and aids in formation of lipid droplets^{15,121}. Demonstrated in Chapter 2, co-expressing *AtWRI1*¹⁻³⁹⁷ and NoLDSP with the soluble squalene pathway increased squalene yields in a *Nicotiana benthamiana* transient expression system, and even further increases were seen when fusing *AtFDPS* and *MaSQS* CΔ17 to NoLDSP, scaffolding the pathway on the surface of lipid droplets.

In this work, we aimed to engineer poplar for squalene production either through cytosolic lipid droplet scaffolding and co-production, which demonstrated the highest squalene yields in transient expression, or in plastids, which is where poplars natively produce isoprene (Figure 3.1). The hybrid poplar NM6 (*P. nigra* L. x *P. maximowiczii* A. Henry) is a female, clonally propagated, commercially valuable clone which has demonstrated rapid growth and biomass accumulation in northern climates¹⁵⁰ and is amenable to engineering through *Agrobacterium*-mediated transformation¹⁵¹⁻¹⁵³. We engineered poplar NM6 with overexpression constructs of squalene pathways, measured yields of transgenic lines throughout the crop, determined the effects of these pathways on photosynthesis and isoprene emission, and performed a technoeconomic analysis to gain insight into future production requirements. We also attempted to engineer the routinely manipulated poplar hybrid P39 (*P. alba* x *P. grandidentata*), though we were unable to generate transformants. While our efforts in poplar P39 were unsuccessful, engineering poplar NM6 to redirect terpenoid precursors away from isoprene and towards terpenes and terpenoids of biotechnological interest may add value to poplar plantations, reduce greenhouse gas emission due to isoprene, and advance poplar as a bioproduct feedstock.

Results and Discussion

Engineering squalene production in transgenic NM6 poplar

To introduce these compartmentalized squalene pathways into hybrid poplar, two constructs for the plastid targeted squalene pathway, three constructs for the cytosolic lipid droplet pathways, and an empty vector were created for transformations (Figure 3.2). A modified pEAQ-*HT* vector⁷² containing two multiple cloning sites (MCS1 and MCS2) was used to generate construct variations for each transformation. The LP4/2A hybrid linkers¹⁰⁴, which enable co- and post- translational cleavage into separate protein products, were used to separate genes in a single MCS to allow expression of multiple genes from a single promoter (Figure 3.2). The two plastid targeting constructs were generated to contain either *CfDXS* and *AtFDPS* in MCS1, separated by LP4/2A, and *MaSQS CΔ17* in MCS2 (pDF1S2), or all three genes in MCS1, each separated by LP4/2A sequences, and an empty MCS2 (pDFS1E2). Three cytosolic lipid droplet constructs were created similarly but using *ElHMGR*¹⁵⁹⁻⁵⁸² and *AtWRI1*¹⁻³⁹⁷ in addition to *AtFDPS* and *MaSQS CΔ17*, with variations in NoLDSP fusion combinations (Figure 3.2). These three constructs were designed to enable co-production of lipid droplets and squalene with either soluble *AtFDP* and *MaSQS CΔ17* (pWL1HFS2), soluble *AtFDPS* and *MaSQS CΔ17* fused to NoLDSP (pWSL1HF2), or both *AtFDPS* and *MaSQS CΔ17* fused to NoLDSP (pWH1SLF2).

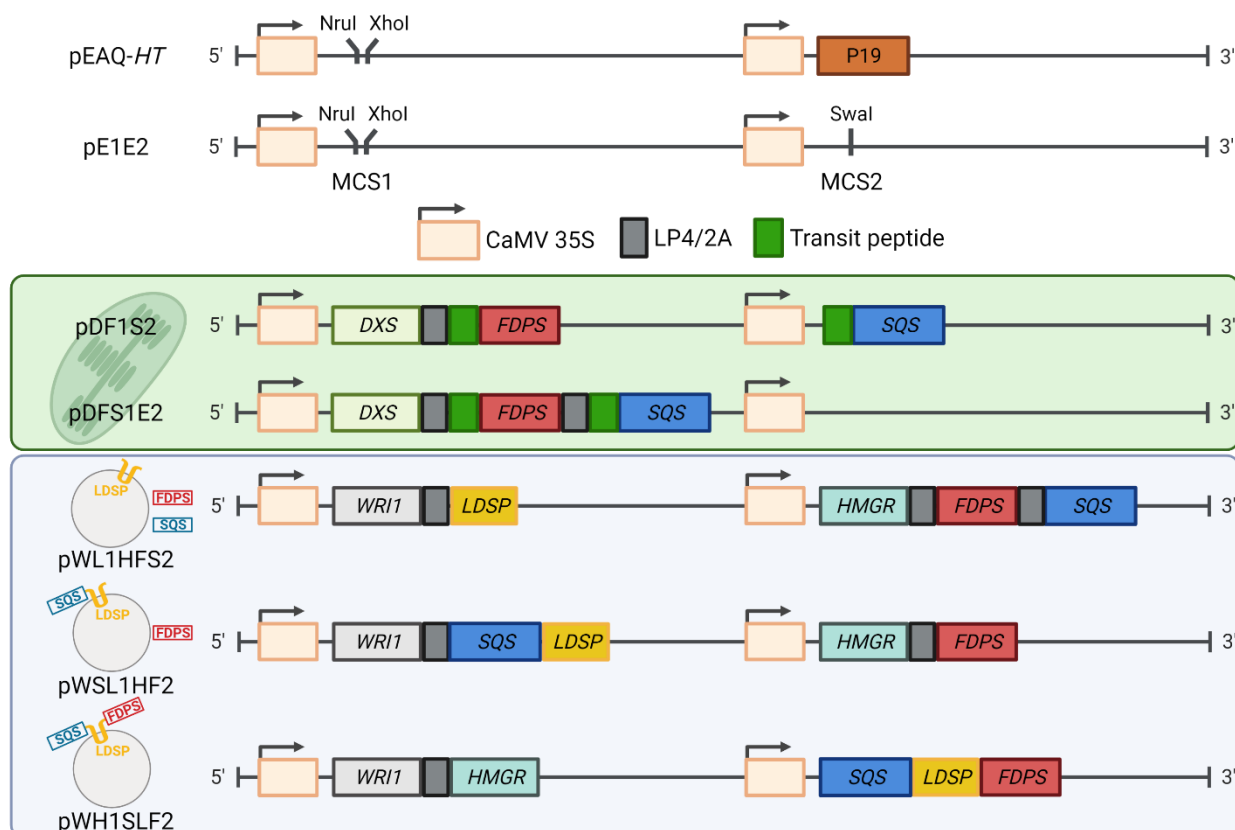


Figure 3.2: Construct design for squalene pathways used for poplar transformations. Constructs used in this study were derived from the pEAQ-*HT* vector (top) modified to contain two multiple cloning sites (MCSs).

Successful transformation of NM6 was initially only achieved with the empty vector construct, pE1E2 (“E” lines), and one of the constructs for plastid targeted squalene production, pDF1S2 (“F” lines). During regeneration of transformed poplar NM6, it was difficult to obtain transformants with pWL1HFS2. In some instances, shoots were formed from callus tissue, but upon transfer to rooting medium plantlets did not survive. To determine whether this was unique to NM6, this construct along with two other constructs with variations of squalene pathway lipid droplet scaffolding (Figure 3.2), were used in attempts to transform the hybrid poplar P39. Poplar P39 is a hybrid line suitable to lab manipulation and is commonly used for generating transformants. When transforming poplar P39, however, similar effects were observed for all three lipid droplet constructs

where plantlets become chlorotic on rooting medium and do not survive, whereas transformed plantlets were able to be generated with the empty vector pE1E2. Further transformation attempts of poplar NM6 eventually led to generation of transgenic plantlets with pWL1HFS2 (“G” lines). However, upon analysis of genomic DNA, only the selection marker located toward the left border was detected, while sequences near the right border were undetectable. Still, these lines were included for analysis to determine if they could produce squalene at higher levels than empty vector lines.

A general analysis of squalene production was performed on leaf tissue across transformants and micro-propagated clones (Figure 3.3). One transformant with the pDF1S2 construct, F4-1, produced the highest squalene yields and the clones, F4-2 and F4-3, produced the next highest yields (Figure 3.3a). While F5-2 was the independent transformant which produced the second highest yields, it was significantly less than F4-1. The two independent lines transformed with pWL1HFS2, G1 and G71, produced detectable levels of squalene, but not more significantly than the empty vectors (Figure 3.3b). Similar mean squalene yields were seen as before (Figure 3.4a) and, in general, higher yields were seen in older leaves (Figure 3.4b). The highest mean yield across leaf stages was 0.63 mg/g fresh weight (mg/gFW) as produced in line F4-1, with the highest mean yield in the 7th leaves at 0.89 mg/gFW. Possible somaclonal variation is seen between micro-propagated clones when comparing squalene yields, especially in F4 lines. Studies have indicated significant changes in DNA methylation levels between early generations of micro-propagated clones, which may be one explanation of phenotypic variation seen here^{154,155}.

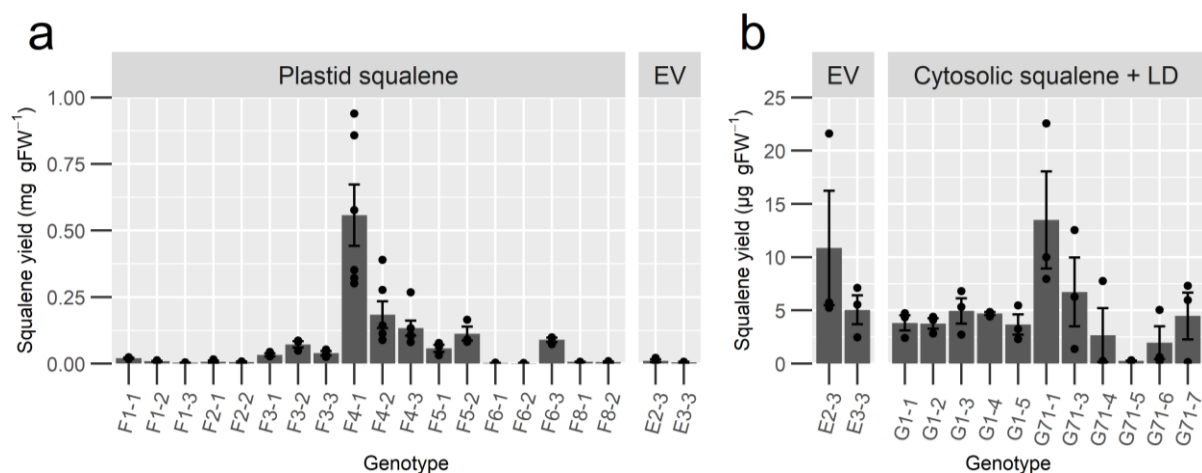


Figure 3.3: Survey of transgenic poplar for production of squalene in leaf tissue. Panel (a) shows poplar lines transformed with the plastid targeted squalene pathway in vector pDF1S2 compared with empty vector (EV) transformed poplar and panel (b) shows poplar lines transformed with the cytosolic and lipid droplet (“LD”) co-production vector pWL1HFS2 compared to the same EV samples from panel (a). Leaf samples were taken at random across six leaves for each indicated transformant, except empty vector lines E2-3 and E3-3 were 3 leaves each. Bars represent the mean squalene yield, with dots representing individual measured leaf samples and error bars representing standard error.

While successful transformants producing squalene were generated for the plastidial squalene pathway, it appears that either the overproduction of cytosolic squalene, lipid droplets, or both may be toxic to regenerating poplar plants, though we were unable to confirm this before the plantlets died. When shoots were regenerated from callus, they would often not survive when transferred to rooting medium, suggesting there may be a root specific regeneration issue when these pathways are expressed. It is likely the lines that were eventually regenerated did not contain the complete construct, considering only the selection marker, but no region near the right border, was detected and no significant production of squalene was measured. Further analysis of young, green stems and root tissue of plastidial squalene lines confirm production is occurring throughout the plants, though with significantly lower yields than leaves (Figure 3.4c and 3.4d).

It can only be hypothesized why the lipid droplet and squalene co-production pathways appear toxic to the plants. Lipid droplet formation was seen transiently expressing *WRI1* and *EYFP-NoLDSP* in poplar NM6 (Figure B.1), confirming the functionality of the platform. It has been demonstrated that *WRI1* can influence auxin homeostasis^{156,157} and auxin has been shown to have an essential role in formation of adventitious roots in species like poplar¹⁵⁸. In this case, overexpression of *WRI1* in stems and roots may be disrupting auxin homeostasis and therefore interfering with adventitious root formation, which is essential for micropropagation of poplar.

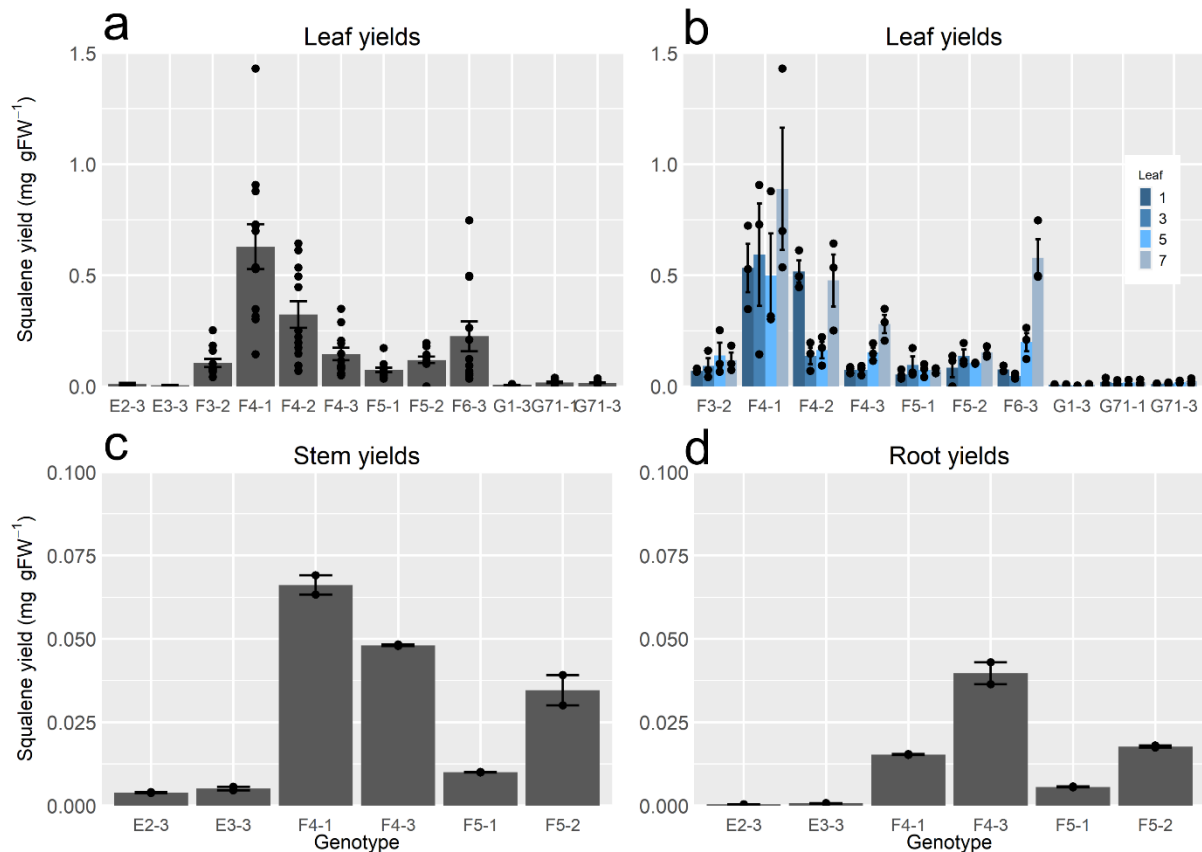


Figure 3.4: Leaf stage, stem, and root analysis of squalene yields in select poplar transformants. The mean squalene yield across four leaf stages (a) and the mean yield at each of these four stages (b) was measured. Samples were collected from the first fully expanded leaf (leaf 1), third, fifth, and seventh leaves from three separate branches, representing three biological replicates, with the mean squalene yield for each stage shown in (b) and the mean squalene yield for all 12 samples in (a). Analysis of squalene

Figure 3.4 (cont'd)

yields in young stems (c), and in roots (d). Empty vector lines E2-3 and E3-3 in panel (a) represent the same data from Figure 3.5. Each stem and root mean were calculated from two replicates of separate extractions from the same bulk tissue. Dots represent individual measurements and error bars represent standard error.

Transient expression in *N. benthamiana* (Chapter 2) demonstrated squalene yields using lipid droplet scaffolding were more than twice that of the plastidial squalene pathway. These results suggest significant increases in transgenic poplar NM6 may be possible through implementation of lipid droplet scaffolding strategies. Future work could incorporate tissue specific expression of lipid droplet and squalene co-production pathways to avoid negative impacts on plant regeneration and development. For example, using leaf specific promoters for each gene in the pathway to reduce possible toxic effects from squalene and lipid droplet co-production in other tissues.

Analysis of isoprene emission and photosynthesis

While Chapter 2 investigated the effects of these engineered pathways on the *N. benthamiana* when transiently expressed, I sought to determine whether there were similar effects on photosynthesis in the transgenic poplar lines. Additionally, with poplar known to emit large amounts of isoprene natively, I aimed to measure how introducing direct competition for IDP/DMADP by the engineered squalene pathway influences total isoprene emission. Three clones of F4 (F4-1, F4-2, and F4-3), two clones of F5 (F5-1 and F5-2), and two independent empty vector lines (E2-3 and E3-3) were chosen to measure changes in isoprene emission and photosynthesis, and how these correlated with differences in squalene yields (Figure 3.5). Compared to the empty vector lines, the five squalene producing plants reduced isoprene emission, with significant reduction in F4-1, (Figure 3.5b) and all except F4-3 significantly reduced photosynthesis (Figure 3.5c).

Analysis of squalene production, isoprene emission, and photosynthesis was repeated, and a significant isoprene emission reduction was seen compared to the empty vector line E2-3, but not E3-3 (Figure B.2). In this experiment, however, squalene yields were reduced across all lines and photosynthesis was not significantly affected, suggesting a need to repeat the analysis. When the plastid targeted squalene pathway was previously tested through transient expression in *N. benthamiana* (in Chapter 2), reduction in photosynthesis was seen. Interestingly, there were no significant differences in photosynthesis in line F4-3, though the other two clones, F4-1 and F4-2, displayed significant reductions. The poplar line with the greatest reduction in isoprene emission was F4-1, which demonstrated the highest squalene yields. Future engineering could incorporate the plastid targeted membrane scaffolding developed in Chapter 2 to potentially reduce the negative effects plastid squalene production has on photosynthesis.

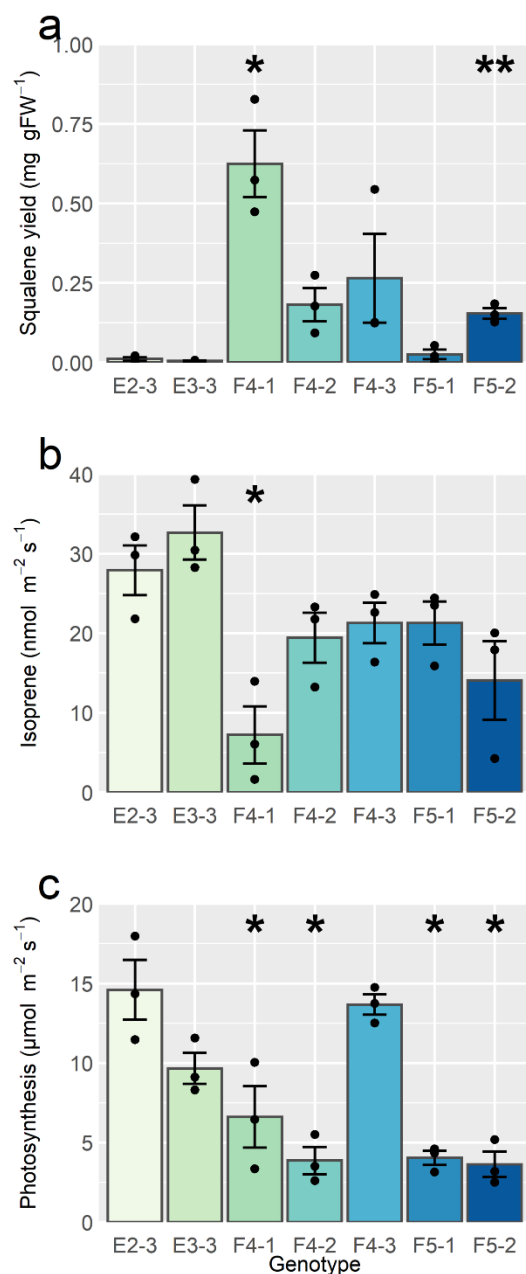


Figure 3.5: Analysis of isoprene emissions and photosynthesis in squalene producing poplar lines. Five plastid squalene producing transformants, three F4 and two F5 clones, and two empty vector lines were chosen to investigate how squalene production (a) may be affecting isoprene emissions (b) and photosynthesis (c). For each plant the seventh fully formed leaf was selected from three branches to measure isoprene emission and photosynthesis with biological triplicates. Following measurements, the measured leaves were collected for squalene extraction. Each dot represents individual measurements and error bars represent the standard error. Asterisks indicate statistical significance compared to E2-3 (*: $p < 0.05$; **: $p < 0.01$) as determined by t test.

Isoprene emission in poplar has been implicated to have roles in various biotic and abiotic stresses^{68,159}, in particular tolerance to heat¹⁶⁰. However, studies have shown non-emitting poplars demonstrate similar biomass productivity to isoprene emitting lines under more temperate conditions^{161,162}. In particular, transgenic lines with suppressed isoprene emission through RNA interference to reduce *ISPS* expression maintained similarly high biomass productivity in a 4-year field trial¹⁶². Additionally, this study found increases in expression of compensatory pathways of protective compounds may have enabled this high productivity. This suggests poplar may be amenable to further engineering of plastid terpenoid pathways to redirect IDP/DMADP away from isoprene and towards squalene and other terpenoids with industrial applications.

Technoeconomic analysis of poplar NM6 squalene production

To begin understanding the economic viability of engineered poplar NM6 for large-scale production of squalene, a technoeconomic analysis (TEA) was performed. A lab-scale, bulk tissue extraction of poplar leaves was performed to simulate squalene extraction and hexane recovery in an industrial setting (Figure 3.6). Analysis was performed assuming only production and extraction from leaf tissue with an average squalene yield of 0.63 mg/gFW. The lab-scale simulation resulted in a hexane loss of 30% due to a lack of separation between leaf tissue and the hexane layer, and this resulted in a minimum selling price (MSP) of squalene to be \$743/kg (Figure 3.6). This is an order of magnitude higher than the current median cost of plant-derived squalene of \$40/kg or shark-derived squalene of \$45.75/kg¹⁶³. Analyzing the TEA results, it is clear the two major limitations in large-scale production would be hexane recovery and feedstock cost, which could both be made up for by increasing squalene yields (Figure

3.6). Performing the TEA with a hexane recovery at 95% reduced the squalene price to \$150/kg, which comes in at 375% of commercial competitiveness. Use of a decanter centrifuge is commonly used in industrial extraction processes, which may allow a much higher hexane recovery, and future TEA simulations could be performed modeling this process. These models do not include microbial production from converted lignocellulosic biomass, as was performed with sorghum⁶³, which could provide further insight into how terpenoid producing poplar may improve economics of the platform. Regardless, it is clear improvements in squalene yields are needed to make this platform economically viable at an industrial scale.

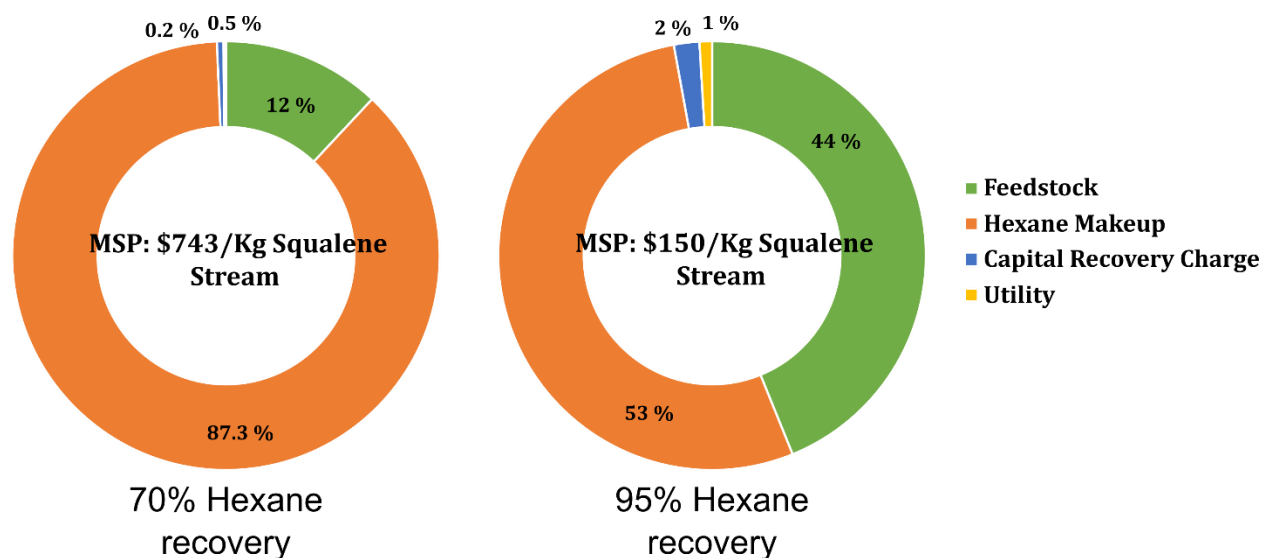


Figure 3.6: Technoeconomic analysis of squalene production, extraction, and purification from poplar leaves. Results from technoeconomic analysis with either 70% hexane recovery as measured in the lab simulation with centrifugation to separate tissue or 95% hexane recovery, which is attainable through separation with a decanter centrifuge. The breakdown of feedstock input and utility costs to extract and purify squalene are presented in Figure B.3 and other assumptions in Table B.1.

Conclusions

In this work, I engineered poplar for production of squalene, which may provide an alternative, sustainable source. Transgenic lines produced up to a mean of 0.63 mg/gFW across leaf stages and accumulated much smaller amounts in stems and roots. While the TEA results indicate that further optimization is needed to improve economic viability, there is great opportunity to further engineer squalene production. To improve plastid targeted squalene production, one strategy which may greatly improve yields would be knockdown or knockout of ISPS to direct more IDP/DMADP towards squalene. As demonstrated in transient expression systems in Chapter 2, the cytosolic lipid droplet scaffolding strategy produced more than twice the amount of squalene compared to the plastidial pathway. Here, transgenic poplar lines could not be generated containing this pathway which may be due to constitutive expression of genes interfering with adventitious root formation. Analysis of root and stem tissue indicates activity of the pathways, which may in-turn cause the developmental issues when regenerating transformants. Future engineering strategies could incorporate tissue-specific expression of desired pathways to avoid any possible unintended effects in other tissues or during regeneration. This work demonstrates novel engineering approaches in a bioenergy feedstock crop, laying the groundwork for a sustainable terpenoid production platform for and as a potential strategy to improve economic viability of such crops.

Methods

Generation of poplar transformants

Constructs for transformation were derived from the pEAQ-*HT* vector⁷² with the P19 gene removed and replaced with a second multiple cloning site (MCS2) (Figure 3.2).

To remove the P19 gene, primers were designed to amplify the entire vector, without the P19 gene, and containing 5' overhangs for re-ligating the vector with a *Sma*I restriction site in place of the gene using In-Fusion cloning mix (Takara Bio). Therefore, both cloning sites were under the control of the constitutively expressing CaMV 35S promoter.

Each set of genes and LP4/2A linkers were consecutively inserted into the indicated MCS (Figure 3.2) using In-Fusion cloning. There are two DNA sequence variants of LP4/2A linkers to avoid identical sequences and assist with cloning. The transgenic poplars were generated as described previously^{152,153}. Briefly, sterile stem internodes or leaves (1 cm long) taken from *in vitro* grown hybrid poplar NM6 were inoculated with *Agrobacterium tumefaciens* strain C58 carrying the selected vector construct. Shoot regeneration was induced in the presence of kanamycin. The regenerated shoots were further screened for antibiotic resistance by transferring to rooting media containing kanamycin. Genomic PCR specific to the selection marker was used to confirm the transgenic events. Genomic DNA was also used in attempts to confirm sequences near the right border in the lipid droplet and cytosolic squalene constructs.

Following rooting and selection marker confirmation, transgenic poplar plantlets were transferred to soil and placed in a growth chamber for acclimation. Chamber conditions were set to a light intensity of 200 $\mu\text{mol m}^{-2} \text{s}^{-1}$ at pot level, 12 hr day length, 23°C during the day and 20°C during the night, and a relative humidity of 60%. Poplar plantlets were allowed to acclimate in the chamber for 2-3 weeks before transplanting to larger pots and transferring to the greenhouse.

Transient expression in poplar NM6

Agrobacterium-mediated transient expression was performed similar to methods in Chapter 2, but in poplar NM6 leaves. *Agrobacterium* strains harboring a pEAQ-HT vector^{72,140} containing either an *EYFP-NoLDSP* fusion or *AtWt1-LP4/2A-EYFP-NoLDSP* were infiltrated with a syringe at an OD₆₀₀ of 1.0. Infiltration was much more difficult than tobacco leaves, however, enough *Agrobacterium* was infiltrated around the infiltration site to measure fluorescence (Figure B.1). Infiltrated leaf areas were cut off and using a razor blade the adaxial layers of leaf tissue were gently scraped off to reveal a thin abaxial layer for imaging in an Olympus Fluoview FV10i microscope. Fluorescent images were taken at excitation:emission wavelengths of 473nm:527nm for EYFP and 559nm:600nm for chlorophyll.

Analysis of squalene production in transformants

Leaf tissue was collected, extracted, and measured using gas chromatography with flame ionization detection (GC-FID) similar to methods in Chapter 2. 15 mm leaf discs were cut, weighed, frozen in liquid nitrogen in 2 mL screw cap tubes containing 0.1 mm glass beads and two 3 mm tungsten carbide beads, and stored at -80°C until extraction. Frozen tissues were ground to a fine powder in a Qiagen TissueLyser at 30m s⁻¹ for 2 min until the tissue was a fine powder. 600µL of hexane containing 50 ng/µL of n-hexacosane was added and samples shaken at room temperature for two hours, 300 µL of water added to aid in separation, and samples were centrifuged at 17,000 x g for 5 mins before transferring the hexane layer to GC vials for analysis. Samples for stem and root extraction were collected from multiple young, green stems or roots, flash frozen in liquid nitrogen and stored at -80°C until extraction. Stem or root samples were ground

with a mortar and pestle into a fine powder, approximately 500 mg of ground tissue weighed in 2 mL screw cap tubes, and squalene extracted as described above. Squalene was quantified as described in Chapter 2.

Analysis of isoprene emission and photosynthesis

Isoprene measurements were recorded in real time using a Fast Isoprene Sensor or FIS (Hills Scientific, Boulder, Colorado). Isoprene reacts with ozone to produce formaldehyde and glyoxal that are electronically excited. When they return to the ground state, green light is emitted and photons are detected by a photomultiplier tube¹⁶⁴. We measured isoprene emission and photosynthesis rates simultaneously using the FIS and the LI-6800 portable gas exchange system (LI-COR Biosciences, Lincoln, NE) respectively. The airflow from the LI-6800 leaf chamber was redirected to the FIS for isoprene measurement. The flow rate in the LI-6800 was set at 500 $\mu\text{mol s}^{-1}$ and the FIS flow rate was set such that it draws sample air from the LI-6800 at 600 sccm (420 $\mu\text{mol s}^{-1}$). A 3.225 ppm isoprene standard from Airgas was used for the FIS calibration. First, we determined the background signal by measuring isoprene levels in the air flowing from the gas chamber. Then the leaf was inserted into the chamber and allowed to equilibrate under the following conditions: light intensity of 1000 $\mu\text{mol m}^{-2} \text{s}^{-1}$ (50% blue light and 50% red light), temperature of 30°C, $[\text{CO}_2]$ of 420 $\mu\text{mol mol}^{-1}$ and water vapor content of 22 mmol mol^{-1} . A measurement was logged after both photosynthesis and isoprene reached steady state at the end of the equilibration period. We subtracted the background signal from each reading of isoprene measurement and calculated mean isoprene emission over the time period it stabilized. Photosynthesis was reported by calculating mean of assimilation rates recorded for 1 min after stabilization. Measurements were done in three

leaves for each genotype. At the end of the day following isoprene emission and photosynthesis measurements two 15 mm leaf discs were collected, weighed, and extracted as described above. Squalene quantification was performed as done in Chapter 2.

Technoeconomic analysis of squalene producing poplar

Lab-scale extraction of squalene from poplar leaf tissue was performed to simulate industrial extraction process. Fresh leaf tissue from transgenic poplar line F4-1 was flash frozen in liquid N₂ then ground to a fine powder. Approximately 3 g of ground tissue was weighed in a 50 mL polypropylene conical tube, 5 mL of hexane containing 50 ng/μL added, samples mixed and allowed to incubate with shaking at room temperature for 2 hr. Tubes were centrifuged at 10,000 x g for 5 mins and as much of the solvent layer as possible was removed with Hamilton syringes to accurately determine the volume of hexane remaining in the leaf tissue layer and unable to be recovered. This extraction resulted in ~70% hexane recovery, which was the value used for the simulation, along with a squalene yield of 0.63 mg/gFW as previously experimentally determined (Figure 3.3).

Based on the experimental data, a process simulation model integrating squalene extraction and separation units was developed in Aspen Plus®, and an approximate MSP for the squalene was determined. The designed process was composed of (1) extraction of squalene from poplar leaves, (2) separation of solvent phase from aqueous phase, (3) recovery of solvent, and (4) purification of squalene. A detailed process flow diagram is depicted in Figure B.3 with other parameters and assumptions outlined in Table B.1. Capital and variable operating costs are estimated from the simulation results based on

squalene flow rate of 100 kg/h. Note that fixed operating cost is not considered here. Equipment costs were estimated using an exponential scaling expression based on equipment size and cost data from Aspen Plus® and the literature¹⁶⁵. The costs for hexane makeup and feedstock are the biggest cost contributors so increasing solvent recovery and squalene yield reduces the MSP of squalene.

Author contributions and acknowledgements

JDB and BRH conceived the study. JDB wrote the manuscript with contributions from AS for isoprene and photosynthesis measurement methods and BK for technoeconomic analysis methods. Pathways and constructs were designed by JDB. Poplar P39 transformation attempts were performed by JDB and FU, with design and discussion guidance from SDM. Poplar NM6 transformations were performed by the Michigan State University Plant Biotechnology Resource and Outreach Center. Squalene analysis in transgenic lines was performed by JDB. Isoprene emission and photosynthesis experiments were performed by AS and supported by TDS. Technoeconomic analyses were designed by JDB, BRH, BK, and CTM, and performed by BK using data collected by JDB. I would like to thank Malik Sankofa for assistance in maintaining transgenic poplar lines and help with a portion of squalene extractions for data presented in Figure 3.3.

CHAPTER 4

High-throughput identification, characterization, and engineering of plant bidirectional promoters

Results from this chapter are being submitted for publication as part of the following manuscript:

Bibik, J. D., Baldermann, A., and Hamberger, B. R. High-throughput identification, characterization, and engineering of plant bidirectional promoters.

Abstract

Plant metabolic engineering quickly becomes complex when aiming to introduce large, multigene biosynthetic pathways. Limitations in gene-stacking to reduce insert sizes and the need for multiple selection markers, as well as genetic tools to precisely regulate expression of several genes makes engineering difficult. In this work, I aimed to develop a set of bidirectional promoters, which drive expression of divergent flanking genes, with a range of expression strengths to enable greater gene regulation in more compact constructs. 34 putative bidirectional promoters from poplar (*Populus trichocarpa*) and *Arabidopsis thaliana* were identified, many of which demonstrated a range of fluorescent reporter production. These bidirectional promoters were also paired with various terminator sequences previously shown to increase gene expression, which further broadened expression regulation. In many cases, bidirectional promoter expression appeared to be context specific depending on the promoter orientation and the associated terminator. Synthetic bidirectional promoters were also created either from newly characterized unidirectional promoters or from bidirectional promoters based on predicted regions of dense transcription factor binding sites. Ultimately, this work establishes a high-throughput pipeline to identify, characterize, and engineer bidirectional promoters to improve plant metabolic engineering.

Introduction

Engineering plants as hosts for bioproduction has become an attractive platform for many chemicals that may not be easily chemically synthesized or biosynthesized in other living hosts. Products with interest for commercial uses can be rather complex and require a large number of genes to build an efficient biosynthetic pathway. Compared to many microbial engineering strategies to express complex pathways, one major limitation in plant engineering is a general lack of genetic tools to effectively regulate large pathways in a plant host¹⁶⁶. Designing sustainable plant production of compounds requires insertion of several genes to integrate the biosynthetic pathway, engineer the precursor pathways, and incorporate pathway compartmentalization strategies. Terpenoids, for example, are the most diverse class of natural products and contribute a dominant fraction of plant specialized metabolism. Production of bioactive terpenoids considered for commercial applications typically requires a variety of modifying enzymes, including multiple cytochromes P450, to functionalize terpene core structures which can result in large biosynthetic pathways. In one example, biosynthesis of the anti-malarial sesquiterpenoid artemisinin involves 5 enzymes, including a cytochrome P450 and other dehydrogenases and reductases, to synthesize the product from farnesyl diphosphate¹⁶⁷. Biosynthesis of paclitaxel is an even more complex example, involving at least 19 enzymes to convert geranylgeranyl diphosphate into the chemotherapeutic diterpenoid²¹. The complexity of many terpenoids makes them extremely difficult or impossible to chemically synthesize, so engineering the biosynthetic pathways in hosts, like plants, may be one of the most promising ways to efficiently produce these valuable chemicals. However, advancements

in plant metabolic engineering are needed to enable more complex regulation of pathways.

Generating transgenic plants using *Agrobacterium*-mediated transformation, or other methods, as described in Chapter 1 becomes difficult when inserting large pathways to yield high levels of desired products. Major limitations for inserting large pathways include time, selection marker availability, transformation efficiencies, and availability of diverse genetic tools to regulate pathways^{74,75}. Consecutive re-transforming is time consuming, as plant transformations can take several months to complete and require separate selection markers each transformation. Co-transformations using multiple genes on separate constructs are inefficient and still require multiple selection markers. Transformation of large, multigene constructs containing all desired genes often results in low transformation efficiency. In all cases, the multiple genes in a pathway require unique promoters to reduce likelihood of unintended gene silencing effects in the host plant^{74,75}. Additionally, products and intermediates of desired pathways can prove toxic to the host^{24,111,112}, requiring specific regulation of which tissues or cell types genes are expressed. Therefore, installation of complex pathways requires strategies to build constructs containing the entire pathway under precise regulation of each gene involved, while keeping constructs compact to reduce size.

Advancements in recombination-based cloning methods have enabled quick assembly of large constructs⁷⁶, but gene expression has still been limited to a few common promoters which lack diverse expression profiles^{74,75}. There have been recent advancements in characterization and engineering promoters for plant gene regulation, though the extent to how effectively they function in broader plant species warrants further

investigation^{82,83}. Many of these promoters have been developed for tunable expression^{84–87}, tissue specificity^{44,46–49,84,168}, and inducibility^{88,169,170}, all of which may aid in overcoming current metabolic engineering limitations. Of particular interest are bidirectional promoters (BDPs), or promoters which can regulate divergent expression of genes on either side of the promoter sequences and enable gene stacking (Figure 4.1).

Bidirectional promoters are commonly used in microbial systems to improve gene stacking in efforts to express a greater number of genes in more compact constructs^{101,102}. An analysis performed from poplar, *Arabidopsis*, and rice revealed occurrence of putative bidirectional promoters throughout plant genomes¹⁷¹. While there have been several studies characterizing natural BDPs^{45,93,172–174}, there have only been a few successful attempts to engineer plant BDPs^{46,49,95,96}, which may be in part due to difficulties in predicting what drives promoter directionality when considering transcription factor (TF) binding sites (TFBSs)¹⁷⁵. BDPs may prove especially useful not only as providing regulatory tools, but also as tools to create compact, synthetic gene clusters. The hybrid LP4/2A linker enables co- and post-translational cleavage between linked genes, creating operon-like constructs of at least three genes as demonstrated in Chapter 2¹⁰⁴. Combining the LP4/2A linker with BDPs could enable expression of at least six genes from a single promoter, reducing the number of promoters needed and reducing construct sizes of large pathways. Additionally, genes associated with BDPs are often co-expressed but can vary in expression strengths¹⁷¹, providing the potential for development of differential expression and improving fine tuning of genes involved in metabolic pathways.

In this work, I developed a high-throughput strategy to identify and characterize BDPs from *Populus trichocarpa* and *Arabidopsis thaliana*. Using publicly available data, I searched genomic sequences for putative BDPs and selected candidates based on reported expression strengths in leaf tissue. I then developed a high-throughput screening method using *Agrobacterium*-mediated transient expression to investigate activities of BDPs in *Nicotiana benthamiana* based on fluorescence reporters. Using predicted TFBSs within functional BDPs, I truncated select BDPs for construction of synthetic variants (sBDPs). I also isolated and characterized select unidirectional promoters (UDPs) to use for construction of additional sBDPs. Finally, I tested select BDPs in combination with previously characterized terminators shown to improve expression in *Nicotiana benthamiana*⁹¹ to introduce additional regulatory control. This strategy presents an efficient platform for identification, characterization, and engineering of BDPs in plants to develop libraries of promoters to improve plant metabolic engineering.

Results and Discussion

Identification and cloning of promoter sequences

Utilizing the Department of Energy's Joint Genome Institute's Phytozome database, putative BDPs were identified from whole transcriptome and genome sequences from *Populus trichocarpa*¹⁷⁶ and *Arabidopsis thaliana*¹⁷⁷. Putative BDPs were initially identified as genomic regions within 1000 nucleotides between transcriptional start sites of two genes divergently expressed in leaf tissue (Figure 4.1a). Further selection of BDPs was based on high specificity of leaf tissue expression, resulting in library of BDPs with a range of leaf expression and specificity (Table C.1). To expand the list of high leaf specificity, selection was expanded to genomic regions within 2000 nucleotides between

transcription start sites. Promoters were selected with a range of expression levels for the reported flanking genes to develop greater expression regulation. Overall, 24 BDPs from *P. trichocarpa* and 10 BDPs from *A. thaliana* were selected for characterization. Additional analysis of putative *Pt*BDPs was performed based on expression data provided by Dr. Shawn Mansfield (University of British Columbia), which was collected at six consecutive, developmental leaf stages (Figure 4.1b, unpublished). This provides a higher resolution analysis of the native activities for each *Pt*BDP across leaf stages in poplar.

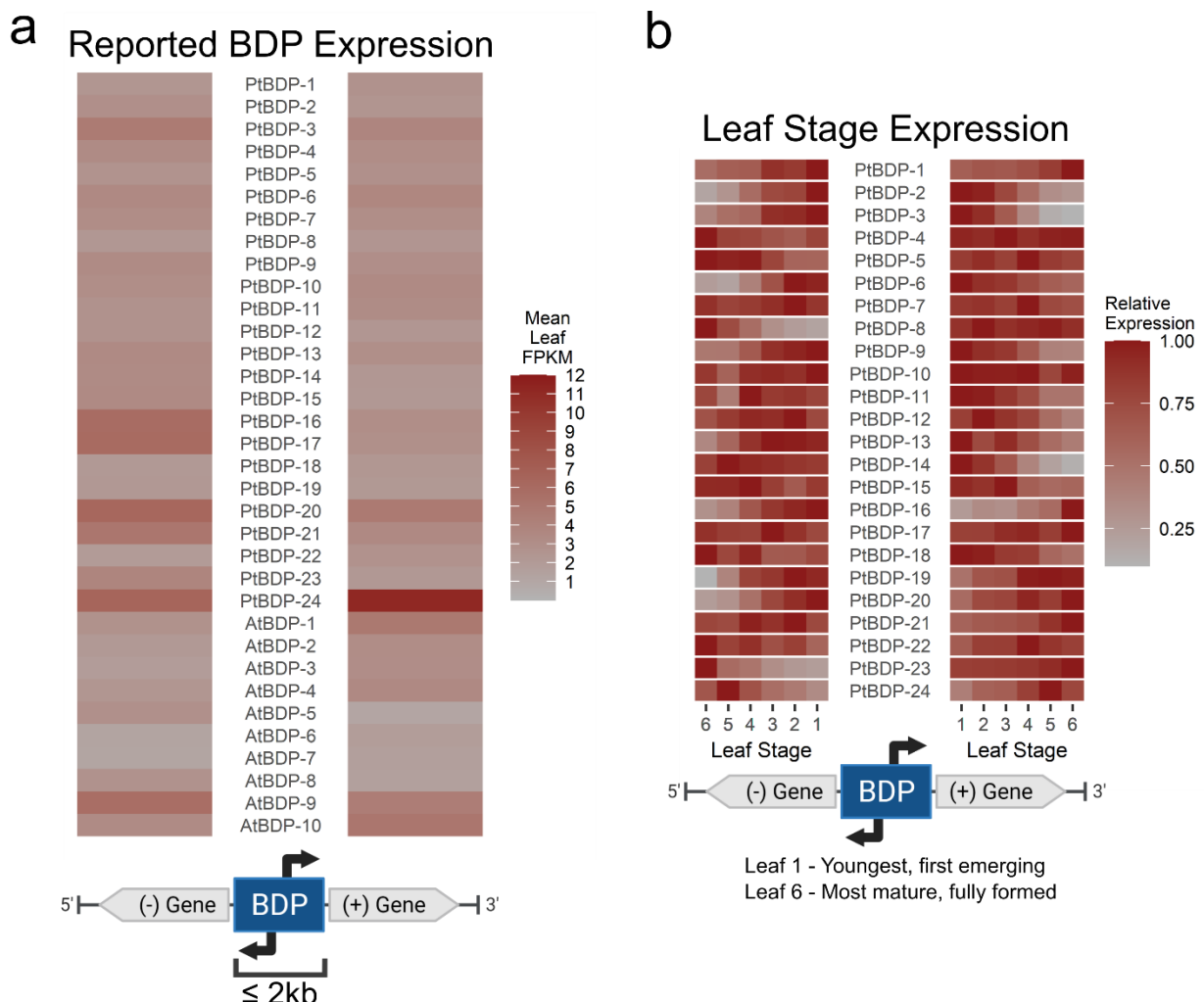


Figure 4.1: Reported leaf expression of selected putative BDPs. Panel (a) indicates the log₂ transformation of the mean Fragments Per Kilobase of transcript per Million mapped reads (FPKM) across reported leaf stages (for *Pt*BDPs) or across leaves under different nitrogen conditions (*At*BDPs) of each putative BDP. Each BDP was selected with

Figure 4.1 (cont'd)

a genomic sequence of less than 2 kilobase pairs (kb) between transcription start sites of flanking genes. Panel (b) indicates relative expression across six consecutive leaf stages, where leaf 1 is the first emerging leaf and leaf 6 is the most mature leaf, based on *P. trichocarpa* expression data provided by Dr. Shawn Mansfield, University of British Columbia (unpublished). The relative expression is based on the leaf stage with the highest FPKM within each expressed gene.

Candidate promoters were amplified along with the predicted 5' UTRs from genomic DNA using primers indicated in Table C.2. Many BDPs had single nucleotide polymorphisms, insertions, or deletions compared to the reference genome. Amplified and verified sequences are given in Supplementary File 1. BDP clones with the fewest genomic differences to the reference sequence were chosen for further analysis. While most of the putative BDPs were cloned, some could not be amplified and subcloned into the pJET1.2 vector or the subsequent reporter vectors. The low G-C content of promoters and 5' UTRs in some instances made design of efficient primers for amplification challenging. Therefore, these BDPs were not analyzed, but the reference genome sequences are included in Supplementary File 1.

Screening promoters in transient expression

To analyze activity of BDP sequences, the pEAQ-*HT* vector⁷² was modified to replace the expression cassette with that of two fluorescent reporters in opposing orientations separated by a cloning site for insertion of BDPs. Two constructs were created containing a cloning site between either mCerulean¹⁷⁸ and Enhanced Yellow Fluorescent Protein (EYFP)¹⁷⁹ (pDualFP) or a Red Fluorescent Protein (TagRFP-T)¹⁸⁰ and EYFP (pDualFP2), with 3' terminator sequences for nopaline synthase terminator (NosT) and mannopine synthase terminator (MasT), respectively (Figure 4.2). Initial testing of pDualFP using the CaMV 35S promoter oriented in either direction showed

fluorescence from EYFP, but no measurable mCerulean fluorescence beyond background levels (data not shown). Therefore, screening of promoters was performed using pDualFP2.

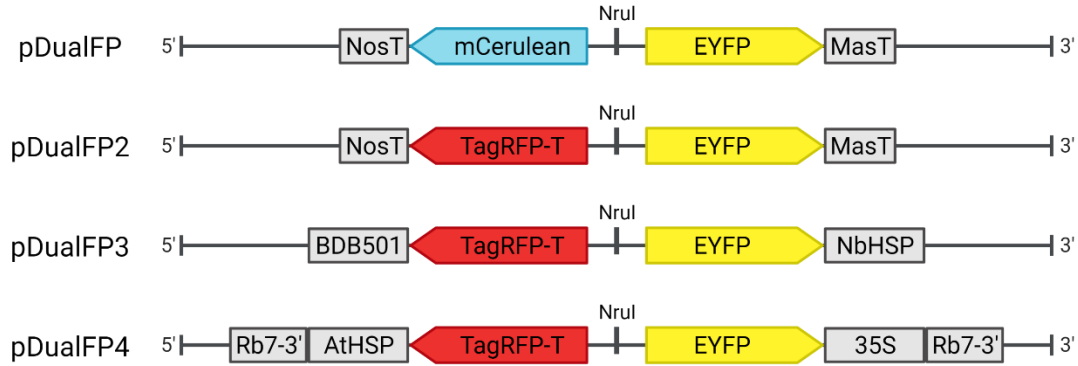


Figure 4.2: Reporter vectors developed for analysis of promoter activities. Each vector was created from the pEAQ-*HT* vector, which also allows for co-expression of the P19 gene to help reduce transgene silencing.

All BDPs were inserted between *TagRFP-T* and *EYFP* in pDualFP2. Each BDP was screened using *Agrobacterium*-mediated transient expression in *N. benthamiana* where crude proteins were extracted, and fluorescence measured. Fluorescence for each BDP is indicated as the relative value compared to the CaMV 35S promoter in the “forward” orientation expressing *EYFP* on the (+)-strand, or in the “reverse” orientation expressing *TagRFP-T* on the (-)-strand (Figure 4.3). Some poplar BDPs (*Pt*BDPs) and *Arabidopsis* BDPS (*At*BDPs) showed little to no activity in these *N. benthamiana* transient expression experiments. Genetic elements can vary in functionality between species so differences of promoter activity in *N. benthamiana* are expected^{181–183}. However, many enabled expression of *TagRFP-T*, *EYFP*, or both with a wide range of measurable fluorescence. Each BDP tested was inserted into pDualFP2 in the “forward” orientation (labelled “F”), which is the 5’ to 3’ sequence as reported in the genome with *TagRFP-T*

and *EYFP* at the 5' and 3' ends, respectively (Figure 4.3a and 4.3c). A select set of *PtBDPs* were also tested in the “reverse” orientation with the 3' and 5' ends inserted near *TagRFP-T* and *EYFP*, respectively (Figure 4.3b).

Most *BDPs* demonstrated *TagRFP-T* or *EYFP* production similar to or less than *CaMV 35S*, but a few notable *BDPs*, *PtBDP16*, *PtBDP24*, and *AtBDP6*, showed increased reporter activity. Comparing fluorescence across *BDPs* and orientations demonstrates not only a range in activities, but also differences in activity based on orientation. For example, some *BDPs*, like *PtBDP13* and *PtBDP18*, showed similar *TagRFP-T* and *EYFP* fluorescence in either orientation, but others like *PtBDP3*, *PtBDP8*, and *PtBDP10* showed a preference for production of *EYFP* regardless of orientation (Figure 4.3a and 4.3b). Across most *BDPs*, *EYFP* fluorescence is more often seen and compared to *TagRFP-T* fluorescence. This may, in part, be due to lower sensitivity of *TagRFP-T* in this system, and use of the more sensitive *superTagRFP-T*¹⁸⁴ may be one way to improve this screening system. However, this does not appear consistently across all *BDPs* tested, particularly when comparing relative fluorescence to *CaMV 35S*. For example, *PtBDP16F* shows nearly 1.5x *EYFP* fluorescence compared to *CaMV 35S* (Figure 4.3a), while *PtBDP16R* only shows 0.65x *TagRFP-T* fluorescence (Figure 4.3b). Therefore, there may be other factors influencing expression based on orientation or other genetic context altering expression of some *BDPs*.

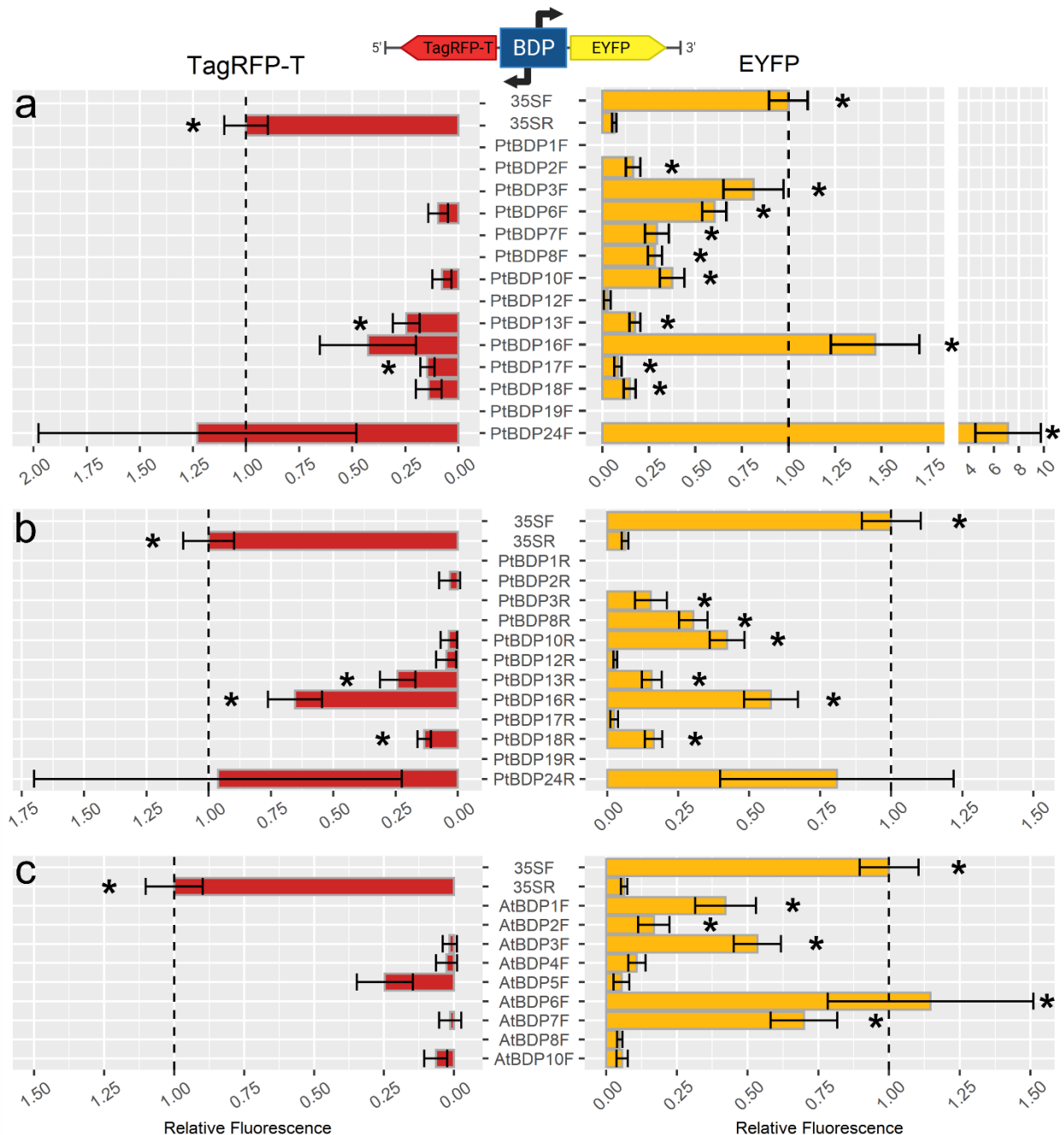


Figure 4.3 (cont'd)

promoter at 1.0, error bars represent standard error, and asterisks indicated statistical significance (p-value < 0.05) compared to the measured background fluorescence from the plants infiltrated with *Agrobacterium* harboring the empty vector pDualFP2.

The BDPs studied here were tested in the heterologous system of *N. benthamiana*, which likely causes the differences or even loss of expression activity from the reported native activity. Additionally, removal of promoters from their native genomic context may remove them from important trans- and cis-regulatory elements that enable the native expression profiles. However, surveying a large number of candidate promoters allowed identification of many BDPs demonstrating a range of activity in *N. benthamiana*. To investigate strategies to further modulate expression strengths using BDPs, *PtBDP3*, *PtBDP10*, *PtBDP 16*, and *PtBDP17* were chosen for further analysis by pairing with previously characterized terminator sequences shown to enhance expression.

Incorporating terminator sequences to further regulate expression

Another recently explored strategy to increase transgene expression in plants is through natural and engineered 3' terminator sequences⁹¹. Not only do terminators regulate effective termination of transcription, but they have also been shown to have a significant role influencing gene silencing of transgenes in plants¹⁸⁵. Damos and Mason (2018) developed a library of 3' sequences which demonstrated a wide range of influence on expression of a fluorescent reporter driven by the CaMV 35S promoter⁹¹. I developed additional pDualFP vectors using select terminator sequences identified by Damos and Mason (2018) to investigate their influence when incorporated into a bidirectional expression system (Figure 4.2). Four terminator sequence variations were chosen to construct the vectors pDualFP3 and pDualFP4 by replacing the NosT and MasT sequences in pDualFP2. The terminators were chosen based on reported expression with

two demonstrating over 5-fold increase in Green Fluorescent Protein production compared to NosT, Bean dwarf mosaic virus movement protein 3' end (BDB501) and *N. benthamiana* Heat Shock Protein terminator (NbHSP). Two synthetic variants demonstrated near 30-fold increase in production, *Arabidopsis thaliana* Heat Shock Protein terminator (AtHSP-Rb7) and CaMV 35S terminator (35S-Rb7), which incorporate a sequence fusion with the Rb7 matrix attachment region (MAR) from tobacco. Due to limitations in DNA synthesis, the Rb7 MAR used in this work is a truncated form, which only contains the last 399 nucleotides at the 3' end (Rb7-3'). To form pDualFP3, the NosT and MasT sequences were replaced by BDB501 and NbHSP, respectively, and for pDualFP4 they were replaced by AtHSP-Rb7-3' and 35S-Rb7-3', respectively (Figure 4.2).

The presence of the terminators had distinct effects on TagRFP-T and EYFP reporter activity between different BDPs (Figure 4.4). For example, AtHSP-Rb7-3' significantly increased TagRFP-T production with *PtBDP16*, but not with *PtBDP3* or the 35S promoter. Expression of *EYFP* using 35S promoter or *PtBDP16* paired with the 35S-Rb7-3' terminator significantly increased production, but this was not seen in *PtBDP3*. It was previously shown that the different terminators can function differently based on the target gene expressed⁹¹. These previous findings in addition to data collected here suggest desired expression of target genes may require an optimal pair of promoter and terminator sequences. This also provides further evidence that BDP activity may change depending on genetic context, including the target gene being expressed. In addition to providing greater regulation of reporter genes here, the terminators improve sensitivity for screening activities of candidate promoters, providing additional tools to further identify

novel promoter activities in future studies. While pairing BDPs with different terminators enabled one strategy to alter expression, I also aimed to engineer select BDPs to further develop expression strengths by altering the BDPs directly.

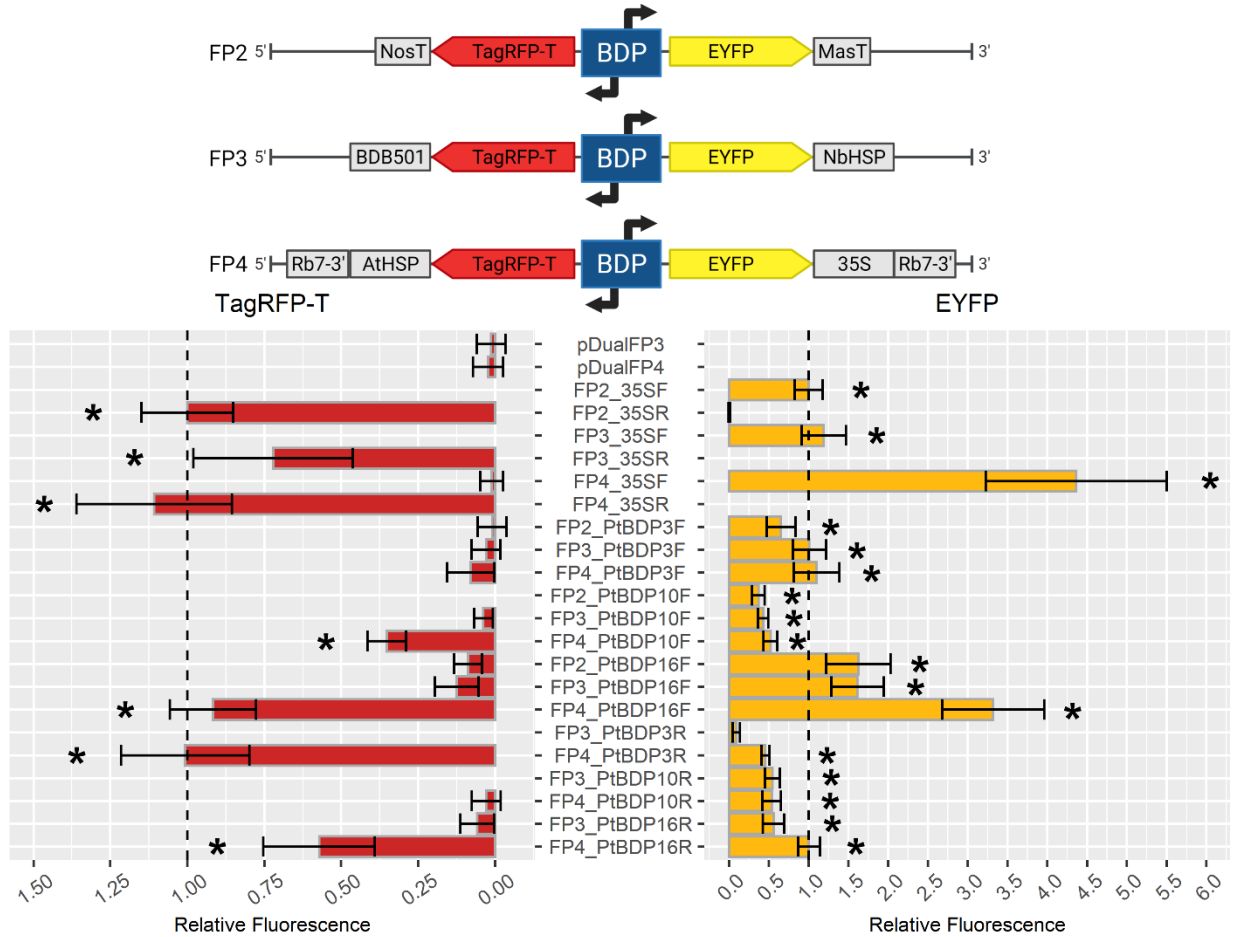


Figure 4.4: Testing select PtBDPs in combination with diverse terminator sequences. The indicated promoters were tested in pDualFP2 (FP2), pDualFP3 (FP3), or pDualFP4 (FP4). Bars represent the mean relative fluorescence compared to the mean TagRFP-T fluorescence (left) with the CaMV 35S promoter in the "reverse" orientation (R) or compared to the mean EYFP fluorescence (right) with the CaMV 35S promoter in the "forward" orientation (F). The mean fluorescence of pDualFP2 infiltrated plants was subtracted from the mean fluorescence of each sample prior to CaMV 35S comparison. Samples in plots with no reported fluorescence have a zero or negative pDualFP2 subtracted mean fluorescence. Dashed lines represent the relative fluorescence from the CaMV 35S promoter at 1.0, error bars represent standard error, and asterisks indicated statistical significance (p-value < 0.05) compared to the measured background fluorescence from the plants infiltrated with *Agrobacterium* harboring the empty vector pDualFP2.

Engineering and characterizing synthetic BDPs

Recent studies have demonstrated the ability to combine elements from different promoters, enabling diversification of expression patterns of engineered promoters^{49,84,87}. Here I engineered sBDPs from existing promoter sequences found in poplar by stacking newly characterized UDPs from poplar head-to-head (5'-to-5') or designing truncated *PtBDP* variants to alter activity and enable creation of novel sBDPs. *PtBDP3* and *PtBDP16* were selected to validate a strategy of predicting regions with abundant TFBSs, truncating around these regions, and building sBDPs with novel expression activities. TFBSs in *PtBDP3* and *PtBDP16* were analyzed using PlantPAN3.0¹⁸⁶ with comparisons to transcription factors found in four dicot species: *P. trichocarpa*, *A. thaliana*, *Glycine max* (soybean), and *Malus domestica* (apple) (Figure 4.5). These predictions were used to guide 5' truncations in areas with a low number of TFBSs and create 3' sequence variations (Figure 4.6). Additionally, three 1kb UDPs from *P. trichocarpa* were selected based on being reported to have high specificity to leaf tissue (Table C.1) and were used to create three initial sBDPs for testing by fusing sequences head-to-head in different combinations (Figure 4.6).

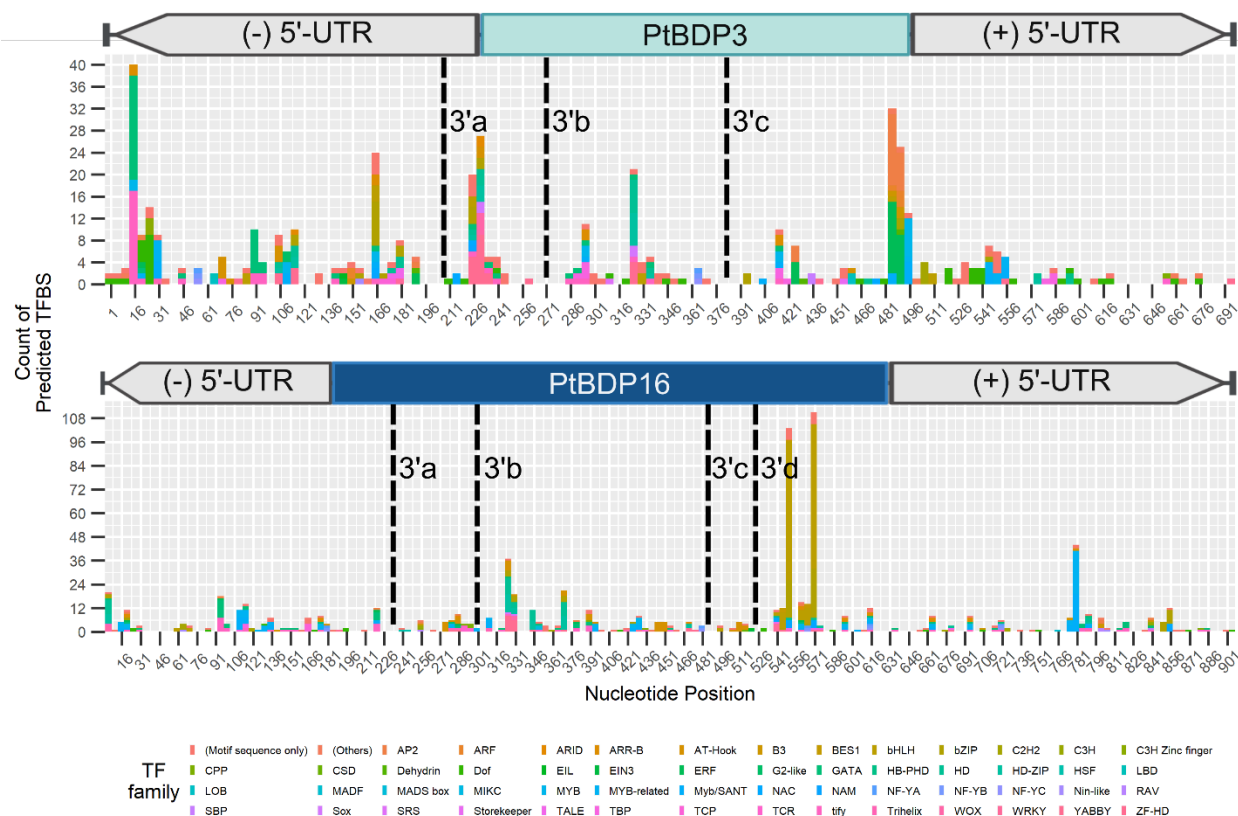


Figure 4.5: Prediction of TFBSs in *PtBDP3* and *PtBDP16* to guide engineering design. The count of TFBSs across the promoter sequences as predicted by PlantPan3.0, binned by every 5 nucleotides. Dashed lines indicate the 5' truncation made to create the 3' sequences indicated. TF family names are those reported from PlantPan3.0 (<http://plantpan.itps.ncku.edu.tw/index.html>).

This approach allowed broad prediction of putative TFBSs to guide sequence selection for activity in *N. benthamiana*. TFBS prediction enabled identification of potentially key regions with abundant TF binding throughout BDPs that may have significant influence on expression (Figure 4.5). Using these predictions, three variations of *PtBDP3* and *PtBDP16* were created by truncating the 5' ends, starting with the (-) – strand 5' UTR, and creating 3' sequences for testing (Figure 4.6). 5' truncations were created because EYFP reporter demonstrated higher fluorescence for both promoters, suggesting greater expression on the (+)-strand (Figure 4.3a and 4.3b). Each truncation was made in regions with no or few predicted TFBSs (Figure 4.5). For *PtBDP3*, the first

truncation was made within the 5' UTR after nucleotide 207 to form *PtBDP3-3'a*⁽²⁰⁸⁻⁷⁰¹⁾ and the second and third truncations made *PtBDP3-3'b*⁽²⁷¹⁻⁷⁰¹⁾ and *PtBDP3-3'c*⁽³⁸¹⁻⁷⁰¹⁾. For *PtBDP16*, truncations were made to form *PtBDP16-3'a*⁽²³⁰⁻⁹¹²⁾, *PtBDP16-3'b*⁽³⁰⁰⁻⁹¹²⁾, and *PtBDP16-3'c*⁽⁴⁸⁶⁻⁹¹²⁾. The promoter variants *PtBDP3-3'b* and *PtBDP16-3'a* are predicted to lack the core promoter region required for (-)-strand gene expression.

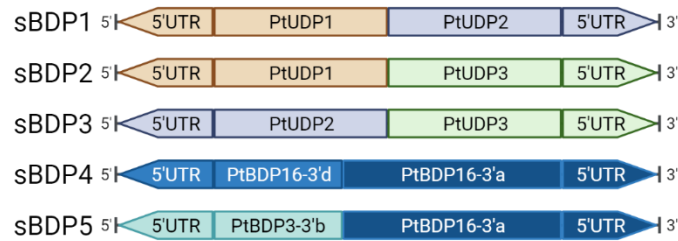


Figure 4.6: Synthetic BDPs created in this study. sBDPs 1, 2, and 3 were created through head-to-head fusions of the indicated *PtUDPs*. sBDP 4 and 5 were created from truncated variants of *PtBDP3* and *PtBDP16*.

These truncated variants were tested through transient expression along the three *PtUDPs* and each sBDP created (Figure 4.7). As a more straightforward approach to develop sBDPs, *PtUDP1*, *PtUDP2*, and *PtUDP3* were used to create sBDP1, sBDP2, and sBDP3 by combining each of them head-to-head (Figure 4.6). The individual *PtUDPs* were screened alongside each sBDP (Figure 4.7a). Similar to other BDPs, these promoters showed a general preference of EYFP production. Each 3' variation of *PtBDP3* and *PtBDP16* demonstrated a range of changes in activity (Figure 4.7b). All truncations maintained activity with significant levels of the (+)-strand EYFP except for *PtBDP3-3'c*. The truncation in *PtBDP16-3'a* appeared to remove activity of the (-)-strand as fluorescence of TagRFP-T was not detectable. Interestingly *PtBDP16-3'a* still maintained similar levels of EYFP fluorescence, indicating expression of the (+)-strand was not

significantly affected. *PtBDP3-3'a* maintained similar EYFP production levels to *PtBDP3* and *PtBDP3-3'b* showed a slight reduction in EYFP.

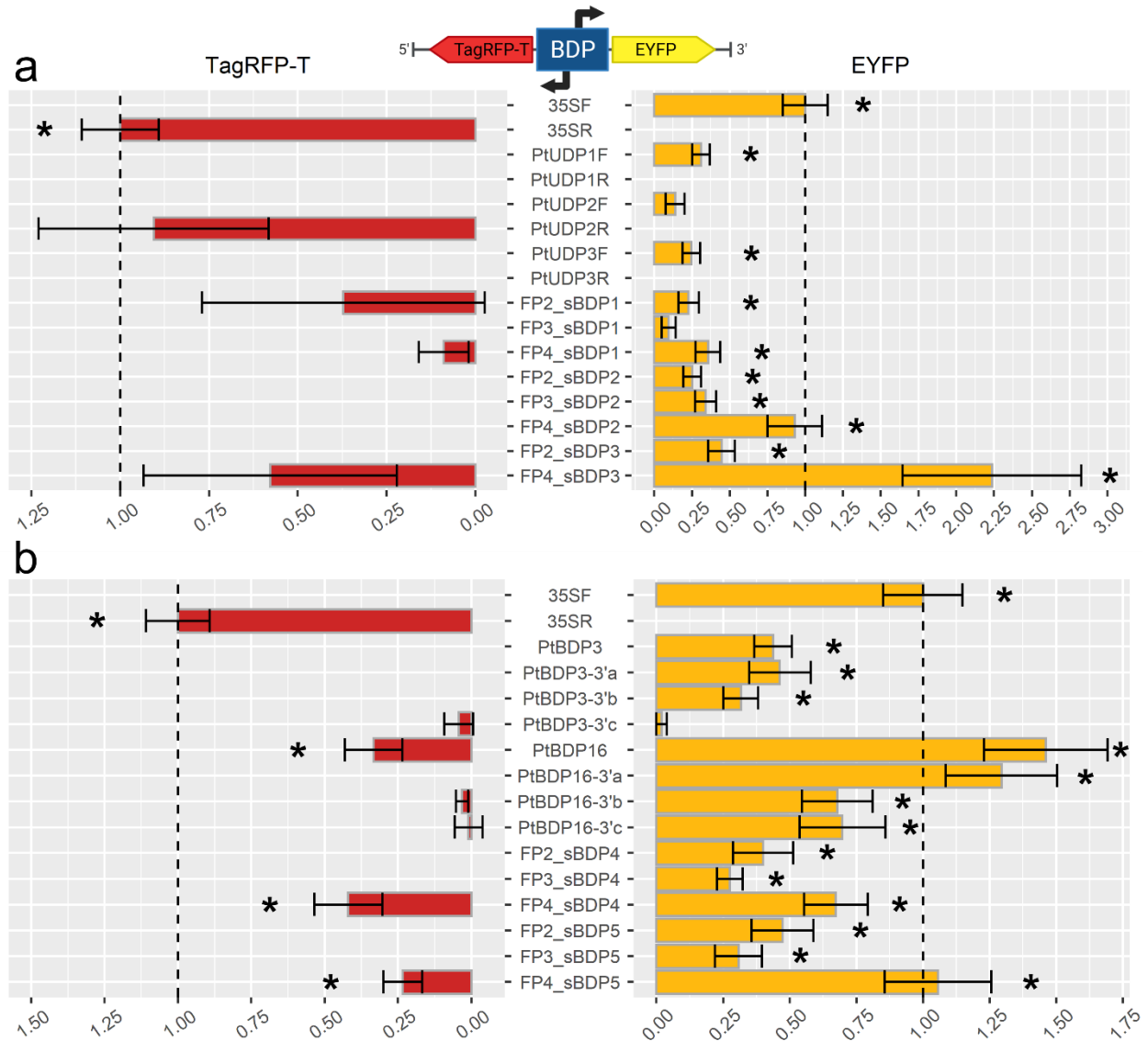


Figure 4.7: Analysis of engineered promoter variants. Panel (a) shows the results of screening sBDP1, sBDP2, sBDP3, and the *PtUDPs* which were used to create them, all in pDualFP2. Panel (b) shows the results of screening 3' sequence variants of *PtBDP3* and *PtBDP16* in pDualFP2 and sBDP4 and sBDP5 in either pDualFP2 (FP2), pDualFP3 (FP3), or pDualFP4 (FP4). Unless indicated by FP2, FP3, or FP4, data represents promoter activity in pDualFP2. The mean fluorescence of pDualFP2 infiltrated plants was subtracted from the mean fluorescence of each sample prior to CaMV 35S comparison. Samples in plots with no reported fluorescence have a zero or negative pDualFP2 subtracted mean fluorescence. Dashed lines represent the relative fluorescence from the CaMV 35S promoter at 1.0, error bars represent standard error, and asterisks indicated statistical significance (p -value < 0.05) compared to the measured background

Figure 4.7 (cont'd)

fluorescence from the plants infiltrated with *Agrobacterium* harboring the empty vector pDualFP2.

The truncated variants *PtBDP3-3'b* and *PtBDP16-3'a* were used to build additional sBDPs (Figure 4.6). For sBDP4, *PtBDP16-3'a* was combined with an additional truncation of *PtBDP16*, *PtBDP16-3'd*⁽⁵²⁴⁻⁹¹²⁾, which contains 104 bp upstream from the transcriptional start site, or the predicted 3' core promoter, and the associated 5' UTR. sBDP5 was created through combination of *PtBDP16-3'a* with *PtBDP3-3'b*. These sBDPs were tested in pDualFP2, pDualFP3, and pDualFP4 to determine activities (Figure 4.7b). Both sBDP4 and sBDP5 demonstrated functionality, but TagRFP-T activity was only detected when in pDualFP4. Creating these sBDPs and combining them with different terminator pairs further demonstrates a strategy to begin developing more fine-tuning of target gene expression, which may enable greater control over individual steps within an engineered biosynthetic pathway.

Conclusions

Here I have developed a high-throughput strategy to identify, characterize, and engineer novel BDPs to further improve plant metabolic engineering. These promoters demonstrate a range of expression strengths, while enabling expression of divergent genes. While many of the putative BDPs showed little to no activity in this *N. benthamiana* transient expression system, those that are functional demonstrate the capability to drive expression of target genes in broader species. Previous studies identifying BDPs have mostly focused in monocot species^{45,46,171,187}, while BDPs characterized here provide a library from two dicot species. Through combination with previously characterized terminator sequences, expression of fluorescent reporters was further altered and

provides a strategy for even greater regulation. Additionally, utilizing TFBS prediction, I was able to design and engineer novel synthetic BDPs. The BDPs described here represent a set of genetic tools that allow for compact gene stacking while enabling differential expression. For example, *PtBDP16* demonstrates high production of reporters which is further amenable to modification, while only being 912 base pairs long with both 5'-UTRs. The BDP library established here provides a set of tools to enable complex metabolic engineering through improvement of gene stacking as well as multigene expression regulation.

Methods

Promoter identification, selection, and cloning

Promoters were identified using genomic and transcriptomic data from *P. trichocarpa* v3.1 and *A. thaliana* Araport 11 obtained from the Phytozome database at <https://phytozome-next.jgi.doe.gov/>¹⁸⁸ (Table C.1). Data files “*Ptrichocarpa_444_v3.1.gene.gff3*” for poplar and “*Athaliana_447_Araport11.gene.gff3*” for *A. thaliana* along with expression data for all genes in either organism downloaded from PhytoMine (<https://phytozome-next.jgi.doe.gov/phytomine/begin.do>) were sorted using R version 4.0.0 with the tidyverse package version 1.3.0. BDPs were identified as genomic regions less than 2000 nucleotides between two divergently expressed genes and most BDPs tested here were selected as regions 1000 or less nucleotides in length. Candidates were further selected based on reported expression of flanking genes in leaf tissue reported as Fragments Per Kilobase of transcript per Million mapped reads (FPKM). The mean FPKM values from all measured leaf stages were transformed to $\log_2(\text{FPKM} + 1)$ where the addition of one to raw FPKM values avoids generating large

negative values when $\log_2$ transforming. The mean transformed FPKM values were calculated across each reported leaf tissue stage or condition. The genes associated with each *PtBDP* tested here had a mean transformed FPKM value of at least 2 across all leaf stages. These values are indicated in Table C.2. *AtBDPs* were selected similarly, but using mean transformed FPKM values across plant treatment conditions and with higher leaf specificity of flanking genes. In this case, *AtBDPs* were selected based on having a mean transformed FPKM values of greater than 1 in leaves and less than 0.5 in roots. *PtUDPs* were also selected for high leaf specificity, where the mean transformed FPKM values for the associated genes were greater than 4 across leaves, less than 2 in stems, and less than 1 in roots.

Genomic DNA from *P. trichocarpa* Nisqually-1 and from *A. thaliana* Col-0 were used as template for amplification using Phusion polymerase (New England Biolabs). Each BDP successfully amplified was inserted into the pJET1.2 vector and sequences were confirmed through Sanger sequencing. Confirmed promoters were amplified from pJET1.2 and inserted into the indicated test vectors, between reporter genes following *Nru*I restriction digest, for screening.

Agrobacterium-mediated transient expression and measurement of fluorescent reporters

Agrobacterium-mediated transient expression was performed similar to Chapter 2. Briefly, *Agrobacterium tumefaciens* strain LBA4404 was transformed with the respective fluorescent reporter construct harboring the intended promoter. Cultures were induced with water and acetosyringone for 2 hours at an OD_{600} of 1.0 prior to infiltrations. For each promoter screening, three leaves on either two or three plants were infiltrated depending

on the experiment, for a total of six or nine replicates, respectively. Plants were allowed to transiently express reporters until samples were collected for analysis on day 4.

Fluorescent protein extraction and measurement was performed similar to a previously developed method¹⁸⁹, but optimized for more high-throughput analysis. On day 4 after infiltrations, four 15 mm leaf discs were collected, placed in 2mL screw cap tubes containing 0.1 mm glass beads and two 3 mm tungsten carbide beads, then flash frozen in liquid nitrogen and stored in -80 °C until extraction. Frozen samples were ground in a Qiagen TissueLyser at 30 m s⁻¹ for 2 mins, allowed to thaw, and ground again. 750 µL of 50 mM sodium phosphate buffer, pH 8.0, was added to samples, mixed thoroughly, and centrifuged at 17,000 x g for 10 min. 150 µL of crude protein extract was added to 150 µL of sodium phosphate buffer in black 96 well plates for analysis of fluorescence. Duplicate plates were prepared. Fluorescence was measured at excitation:emission wavelengths of 550nm:593nm for TagRFP-T measurements and 497nm:540nm for EYFP measurements. Measurements were performed in a BioTek Synergy H1 microplate reader.

Author contributions and acknowledgements

JDB and BRH conceived the study. JDB designed the study, wrote the manuscript, identified, cloned, engineered, and screened promoters. AB assisted with cloning and screening promoters from *Arabidopsis*. We would like to thank Trine Andersen and Malik Sankofa for assistance in maintaining *N. benthamiana* plants used throughout this study. I would also like to thank Dr. Shawn Mansfield for providing developmental leaf stage transcriptomes for *P. trichocarpa*. BDP and vector graphics were created using BioRender.com.

CHAPTER 5

Conclusions and future directions

Summary

Plants are becoming a more desirable engineering platform for production of specialty chemicals and bioproducts. The research performed throughout this dissertation demonstrates strategies to further advance the capabilities of complex metabolic engineering in plants for production of terpenoids. Through pathway optimization and compartmentalization (Figure 2.1), production of squalene was improved while enabling co-production of lipids which may also be of biotechnological importance. Lipid droplet scaffolding of the biosynthetic pathways results in squalene yields over twice as high as the plastidial targeting, but concurrently reduces photosynthesis in the plant (Figure 2.6). However, targeting the scaffolding to plastids ameliorates some of the negative effects on photosynthesis (Figure 2.6). The NoLDSP scaffolding also demonstrates modularity, as different fusion variations result in improved yields (Figure 2.3). These strategies were developed in an *Agrobacterium*-mediated transient expression system and installing them in poplar provides the opportunity to analyze their functionality in a stably engineered host.

Here I show that poplar can be engineered as a platform for production of squalene by generating transgenic lines with plastidial targeted biosynthesis (Figure 3.4). To the best of my knowledge, this is the first case of poplars being engineered for production of terpenoids, and these results establish proof-of-concept towards economic viability on large scale. Given the cytosolic lipid droplet scaffolding demonstrated over twice the squalene yields of plastidial targeting (Figure 2.6), implementing this platform in poplar could significantly improve squalene yields. This pathway appears to have a negative impact during tissue regeneration, however, and I hypothesize this may be due to

overexpression of *AtWRI1*, or the combination with the squalene pathway, interfering with adventitious root development. Squalene production is seen in stem and root tissue with the plastid targeted lines, indicating the pathways are expressed at the junction of adventitious root formation (Figure 3.4c and 3.4d). Additionally, transient expression in poplar NM6 demonstrates co-expressing *AtWRI1* and *NoLDSP* leads to lipid droplet formation (Figure B.1), suggesting the scaffolding strategy is active in transgenic plantlets. Future iterations would likely benefit from regulating tissue or temporal specificity of these pathways to develop transgenic poplar lines.

Having a diverse set of genetic tools available is essential for metabolic engineering of complex pathways. BDPs naturally provide more efficient gene stacking than regularly used UDPs, with many showing different levels of expression for the divergently expressed genes. Building on this idea, the library of BDPs I have developed here provide additional tools to enable more regulation of compact, multigene constructs. Rapid development of these tools is often difficult due to plant transformation limitations but by leveraging transient expression methods here, I could perform rapid identification (Figure 4.1), characterization (Figure 4.3), and engineering (Figure 4.7) of BDPs showing a range of expression. Additionally, combining these BDPs with previously characterized terminator sequences⁹¹ introduced further modularity to regulating expression (Figure 4.4 and 4.7). These provide not only a new set of genetic tools for metabolic engineering, but also a diverse set of sequences which may provide further insight into which regulatory elements may be important for future construction of synthetic BDPs.

Future work

Developing compartmentalization of pathways

Results from Chapter 2 and previous work performed in the Hamberger lab¹⁵ demonstrates scaffolding terpenoid biosynthesis on the surface of lipid droplets is an effective strategy to improve yields and store target compounds. This work was developed for production of squalene, which has commercial uses in cosmetic oils, vaccine adjuvants, and with potential as a biofuel^{107,108,110}. Squalene is also the precursor to an array of higher-value triterpenoids which may also be compatible with the compartmentalization strategies developed here. These strategies can be leveraged for squalene derivatives like ambrein, which is a valuable compound in the fragrance industry^{19,20,109}. Previous work has established ambrein biosynthesis using two bacterial squalene-hopene cyclases, but there are multiple co-products as each enzyme can cyclize squalene and use either product as substrates.^{19,20} This pathway could be used to investigate the capabilities of using lipid droplet scaffolding to drive substrate channeling as synthetic metabolons¹⁹⁰.

Utilizing lipid droplet scaffolding in species already well established for lipid production is also an interesting consideration. Many plant sources have become a target for lipid production¹⁹¹ and scaffolding other proteins important for lipid production or metabolic pathways for co-products may be able to further modulate lipid and lipid droplet formation or add value to production hosts. In addition to plants, microalgae are commonly employed for lipid production^{119,120}, including *Nannochloropsis* where NoLDSP was discovered¹²¹. *Nannochloropsis* could be engineered with a more functionalized LDSP to develop lipid-terpenoid co-production strategies. The algae *Haematococcus pluvialis* has

attracted interest as a platform for co-production of triacylglycerols and the tetraterpenoid astaxanthin^{119,120}. Production of astaxanthin establishes a metabolic capacity for terpenoid production and adapting LDSP scaffolding for squalene and other terpenoids could expand the production capabilities of this alga.

Developing poplar as a terpenoid production host

Research in this dissertation lays the groundwork to utilize poplar as a production host for terpenoids and doing so could add value to the lignocellulosic conversion pipeline and further establish poplar as a bioenergy crop. Plastid targeted squalene production was successful in poplar, yet it caused concurrent reductions in photosynthesis as seen in transient expression experiments. Implementing membrane scaffolding in poplar may ameliorate some of these effects. Previous TEA performed in sorghum predicts engineering bioenergy crops for terpenoid production may reduce the costs of products formed through lignocellulosic biomass conversion⁶³. Performing a similar analysis with the squalene producing poplar model could determine if this is an economically viable option for a crop like poplar. However, the current TEA analysis suggests more engineering is needed to establish poplar as a standalone production platform for squalene, with squalene yields still being a major limitation.

One strategy which could improve plastid targeted squalene yields is by creating isoprene synthase knockouts in transformed lines. While squalene production reduced isoprene emissions, isoprene was still produced even in the highest squalene producing line. As a direct competitor for IDP/DMADP, knockouts of *ISPS* could be the most direct way to further improve flux towards squalene biosynthesis. Given the diverse roles of isoprene in plants^{68,159,160}, however, further investigation would be needed to determine

how these squalene producing, *ISPS* knockout lines would perform in the stress conditions seen in production fields. Implementing lipid droplet scaffolding strategies may also significantly improve yields, but the negative impacts during regeneration prevented transformant generation. It is also possible overproduction of squalene may alter levels of the squalene-derived sterols, many of which are essential for normal plant development¹⁹². However, previous work (in Chapter 2) demonstrated overexpression of *HMGR* alone induces high accumulation of squalene, and a separate study generated *HMGR* overexpression lines in a different poplar hybrid without the same lethality seen here¹⁹³. The lipid droplets may also be sequestering squalene and related metabolites, preventing their conversion towards triterpenoids and sterols important for development. These theories could be further investigated in future iterations of engineered poplar, by inserting genes for lipid droplet production alone and cytosolic squalene production alone. Hypothesizing the lipid droplet and squalene co-production pathways are disrupting adventitious root formation from stems, targeted gene expression could reduce or avoid regeneration interference by using leaf specific or inducible promoters^{83,84,88,194}.

Engineering and expanding available plant BDPs

Development of novel genetic tools will drive complex metabolic engineering in plants. Using transient expression and dual fluorescent reporter vectors established here enabled rapid characterization and engineering of BDPs. While the reporters used here were effective, modifying TagRFP-T to superTagRFP-T¹⁸⁴ may improve the sensitivity and allow more accurate screening of BDPs driving low expression. Results in Figure 4.3a and 4.3b suggest TagRFP-T and EYFP may have different effects on expression strength of promoters. A similar effect was seen in previous work characterizing

terminator sequences, where some terminators showed antagonistic effects when paired with GFP⁹¹. Developing these BDPs in a *N. benthamiana* transient expression system enables rapid testing of promoters and gene combination compatibility. The library of BDPs characterized here also provide a set of sequences for future analysis to guide further engineering of sBDPs. Understanding which elements of the promoters significantly contribute to expression profiles will enable modifying natural BDPs while also providing sequences required to build sBDPs.

The BDPs characterized here could be used to drive expression of terpenoid pathways in more compact constructs. Using these BDPs, expression of lipid droplet scaffolding for squalene production would only require a single BDP, as each side of the BDP could drive expression of two or three genes separated by LP4/2A linkers. For example, *PtBDP16* could express *AtWRI1*-LP4/2A- *EIHMGR*¹⁵⁹⁻⁵⁸² on the (-) – strand while also expressing *MaSQS CΔ17*-LP4/2A-*AtFDPS* on the (+) – strand, a pathway that required four plasmids in transient expression or two MCSs in poplar constructs. The use of these BDPs would be especially effective for much more complex pathways like paclitaxel which is predicted to contain 19 enzymes²¹.

Conclusion

This dissertation presents strategies to improve production of terpenoids in plants as well as tools to facilitate complex engineering of plants with the required pathways. Optimization, re-targeting, and compartmentalization of terpenoid biosynthesis was shown to improve yields and ameliorate some of the negative effects the pathways can have on the host. Introducing these pathways in commercial species may provide sustainable strategies for production of terpenoids, and consideration of tissue specificity

or temporal expression may be required depending on the pathway. BDPs are an effective strategy to improve gene stacking and modulate gene expression, and the natural and synthetic promoters developed here can facilitate engineering of complex pathways for terpenoids and other bioproducts.

APPENDICES

APPENDIX A

Supplemental Data for Chapter 2

Table A.1: Genes used in Chapter 2 and their associated accession numbers.

Gene abbreviation	Organism	Accession number
DXS/nDXS variants		
<i>CfDXS</i>	<i>Coleus forskohlii</i>	KP889115.1
<i>PtDXS</i>	<i>Populus trichocarpa</i>	EU693019.1
<i>RibB G108S (nDXS)</i>	<i>Escherichia coli</i>	WP_139956769.1
<i>YajO (nDXS)</i>	<i>Escherichia coli</i>	NP_414953.2
<i>DXR</i>	<i>Escherichia coli</i>	NP_414715.1
FDP synthases		
<i>AtFDP5</i>	<i>Arabidopsis thaliana</i>	NM_117823.4
<i>PaFDP5</i>	<i>Picea abies</i>	EU432049.1
<i>GgFDP5</i>	<i>Gallus gallus</i>	XM_015298647.1
Squalene synthases		
<i>AhSQS</i>	<i>Amaranthus hybridus</i>	AB691229.1
<i>BbSQS</i>	<i>Botryococcus braunii</i>	KT388100.1
<i>ElSQS</i>	<i>Euphorbia lathyris</i>	JQ694152.1
<i>GlSQS</i>	<i>Ganoderma lucidum</i>	DQ494674.1
<i>MaSQS</i>	<i>Mortierella alpina</i>	KT318395.1
<i>ERG9</i>	<i>Saccharomyces cerevisiae</i>	NP_012060.1
<i>HoDISQS</i>	<i>Haslea ostrearia</i>	AYV97147.1
Other genes		
<i>CasS</i>	<i>Daphne genkwa</i>	MZ485349.1
<i>CfGGDPS</i>	<i>Coleus forskohlii</i>	KP889114.1
<i>AtWRI1¹⁻³⁹⁷</i>	<i>Arabidopsis thaliana</i>	AY254038.2
<i>NoLDSP</i>	<i>Nannochloropsis oceanica</i>	JQ268559.1
<i>ElHMGR¹⁵⁹⁻⁵⁸²</i>	<i>Euphorbia lathyris</i>	JQ694150.1
<i>AtIPK</i>	<i>Arabidopsis thaliana</i>	AY150412.1
<i>RcMPD</i>	<i>Roseiflexus castenholzii</i>	ABU57050.1
<i>AtBCCP1</i>	<i>Arabidopsis thaliana</i>	NM_121644.4

Table A.2: Analysis of photosynthesis response to CO₂ in leaves expressing plastid targeted and cytosolic squalene pathways, with and without NoLDSP scaffolding. *A/C_i* curves were fitted by the Farquhar-von Caemmerer-Berry biochemical model of photosynthesis using the following software: *A/C_i* curve fitting utility version 2.9 for tobacco^{143–145}, to determine how the biochemical capacities underlying photosynthesis: maximum carboxylation rate (V_{cmax}), maximum rate of electron transport (J), triose phosphate utilization rate (TPU), were affected in leaves expressing plastid targeted and cytosolic squalene pathways, with and without NoLDSP scaffolding. Values represent means $\pm$ standard error. $n = 4$ plants per treatment.

Treatment	Before Infiltration			3 Days After Infiltration			5 Days After Infiltration		
	V_{cmax} ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	J ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	TPU ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	V_{cmax} ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	J ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	TPU ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	V_{cmax} ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	J ($\mu\text{mol m}^{-2} \text{s}^{-1}$)	TPU ($\mu\text{mol m}^{-2} \text{s}^{-1}$)
Empty Vector	77.4 $\pm$ 1.8	129.1 $\pm$ 4.7	7.9 $\pm$ 0.3	53.3 $\pm$ 4.9	91.0 $\pm$ 5.8	6.5 $\pm$ 0.3	55.3 $\pm$ 3.2	92.5 $\pm$ 5.6	6.2 $\pm$ 0.3
plast:SQ (-) LDSP	79.9 $\pm$ 2.3	120.2 $\pm$ 9.2	7.3 $\pm$ 0.5	18.4 $\pm$ 1.8	33.4 $\pm$ 2.4	2.3 $\pm$ 0.2	4.1 $\pm$ 0.9	10.0 $\pm$ 1.5	0.7 $\pm$ 0.1
plast:SQ (+) LDSP	76.1 $\pm$ 3.5	118.7 $\pm$ 7.8	7.6 $\pm$ 0.3	42.9 $\pm$ 3.8	68.1 $\pm$ 5.8	4.6 $\pm$ 0.4	15.8 $\pm$ 4.0	30.8 $\pm$ 6.3	2.2 $\pm$ 0.4
cyt:SQ (-) LDSP	80.9 $\pm$ 6.5	118.0 $\pm$ 6.5	7.3 $\pm$ 0.3	46.0 $\pm$ 6.2	72.3 $\pm$ 7.7	5.0 $\pm$ 0.6	30.1 $\pm$ 10.2	56.2 $\pm$ 14.2	4.0 $\pm$ 1.1
cyt:SQ (+) LDSP	82.3 $\pm$ 2.6	124.4 $\pm$ 6.9	7.8 $\pm$ 0.3	32.5 $\pm$ 5.9	57.0 $\pm$ 4.8	3.9 $\pm$ 0.3	16.8 $\pm$ 7.4	34.8 $\pm$ 10.9	2.5 $\pm$ 0.8

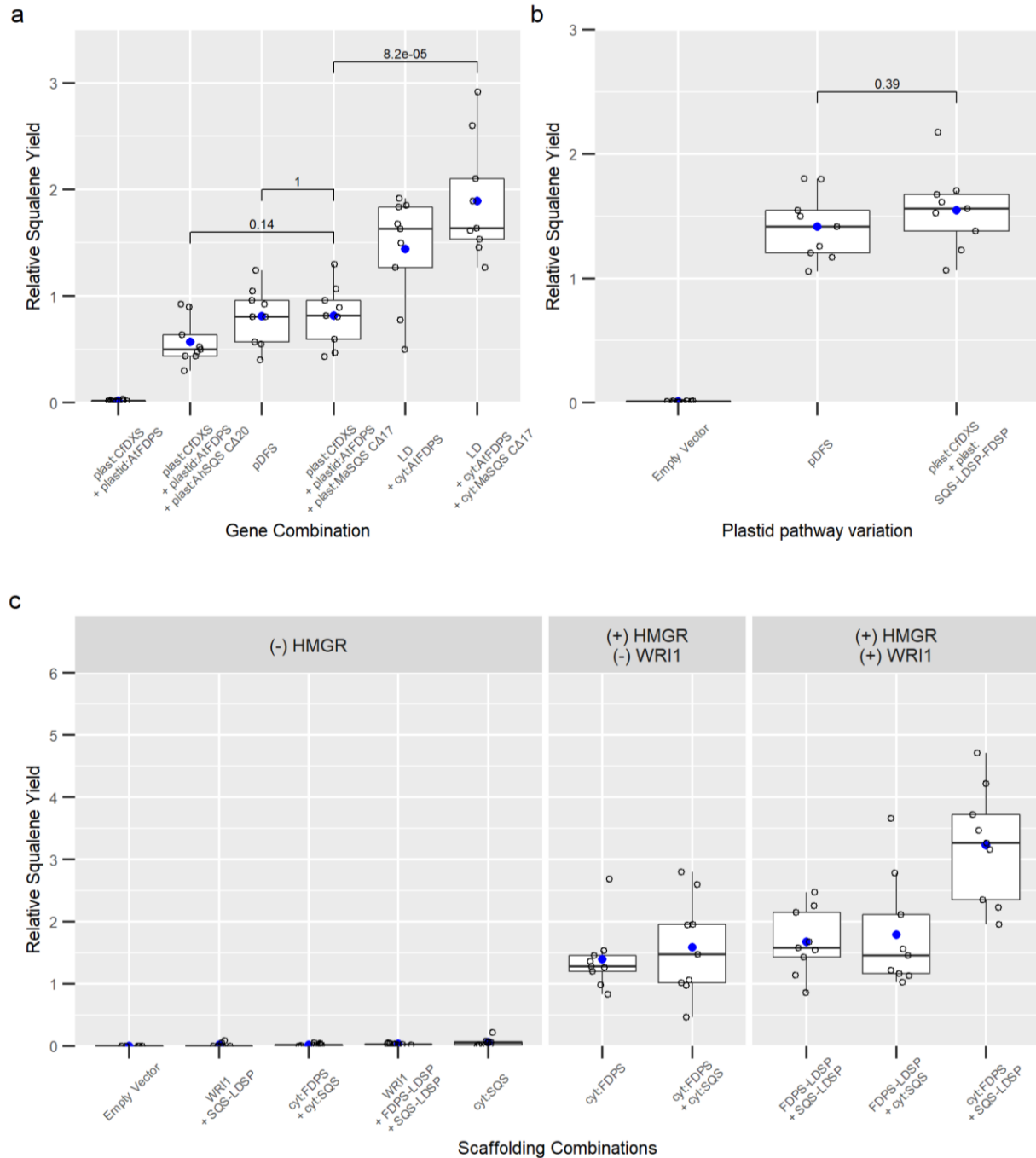


Figure A.1: Additional boxplots comparing soluble and NoLDSP scaffolding pathways in the cytosol and plastids. Panel (a) shows additional plastidial, SQS comparisons with pDFS (plast:*CfDXS*, plast:*AtFDPS*, and plast:*MaSQS C417* separated by two LP4/2A hybrid linkers in pEAQ-*HT*) and the cytosolic co-production of lipid droplets, without scaffolding, and cyt:*EIHMGR*¹⁵⁹⁻⁵⁸², cyt:*AtFDPS*, and cyt:*MaSQS C417*. LD indicates co-expression with *AtWRI1*¹⁻³⁹⁷ and *NoLDSP*. Panel (b) shows comparisons between the pDFS vector and vectors used for plastid scaffolding in the photosynthesis experiments. Panel (c) shows additional combinations of cytosolic lipid droplet

Figure A.1 (cont'd)

scaffolding, including the initially tested *AtFDPS-NoLDSP* fusion variant. Each panel represents data from separate transient expression experiments. Open circles are individual data points, blue circles are mean value, and horizontal line within box represents the median value. The box shows the range from the lower 25th percentile to the upper 75th percentile. The upper and lower whiskers extend to the largest and smallest data point no further than 1.5x the inter-quartile range, with points lying outside the whiskers considered outliers. Paired statistical comparisons were performed by *t* test indicated by brackets with the corresponding *p*-values.

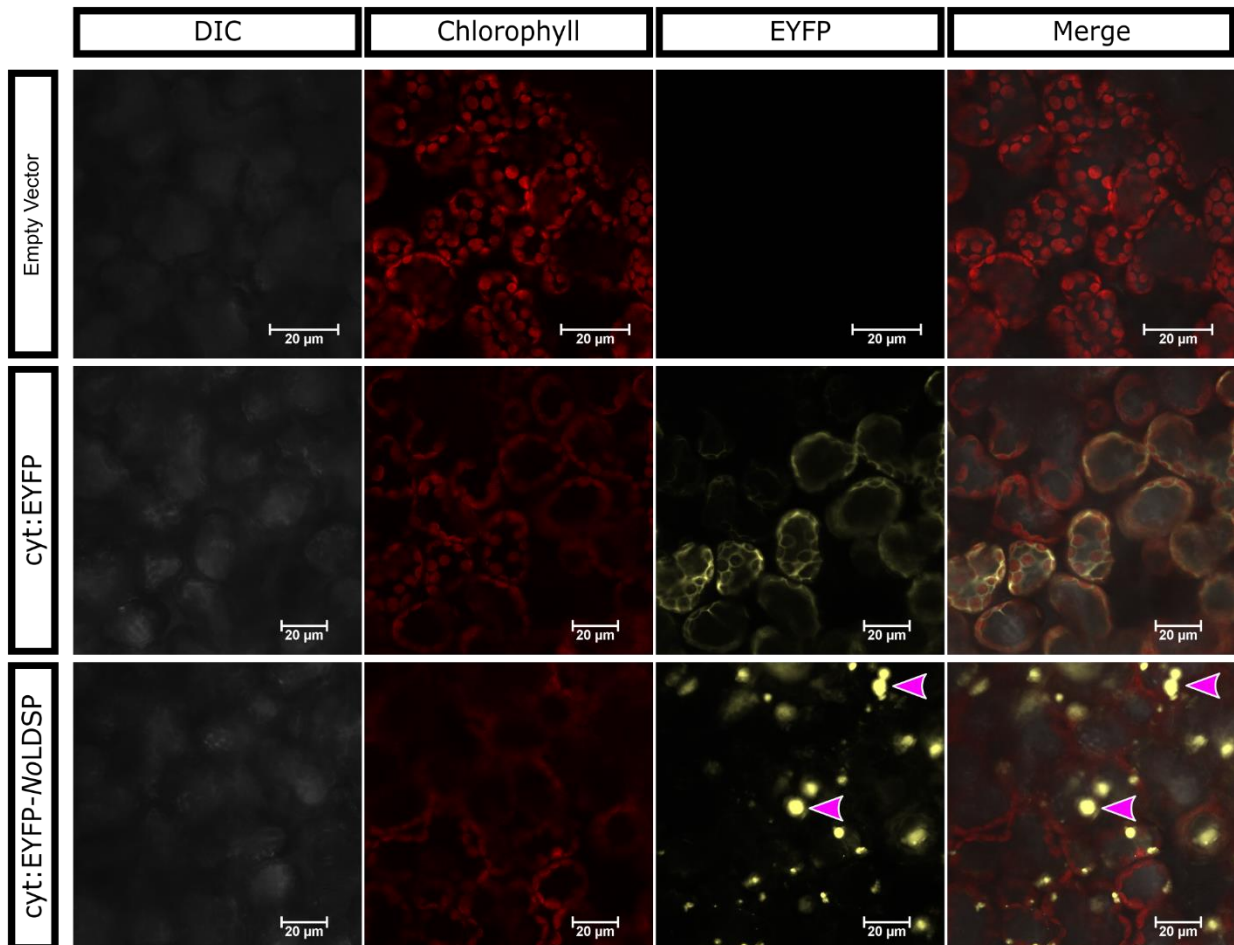


Figure A.2: Confocal microscopy comparing cytosolic EYFP and EYFP-NoLDSP. EYFP fluorescence and chlorophyll autofluorescence were measured with excitation:emission wavelengths of 513.9 nm:585 nm and 561 nm:700 nm, respectively. Pink arrows point to EYFP seen aggregating at lipid droplets in the cytosol.

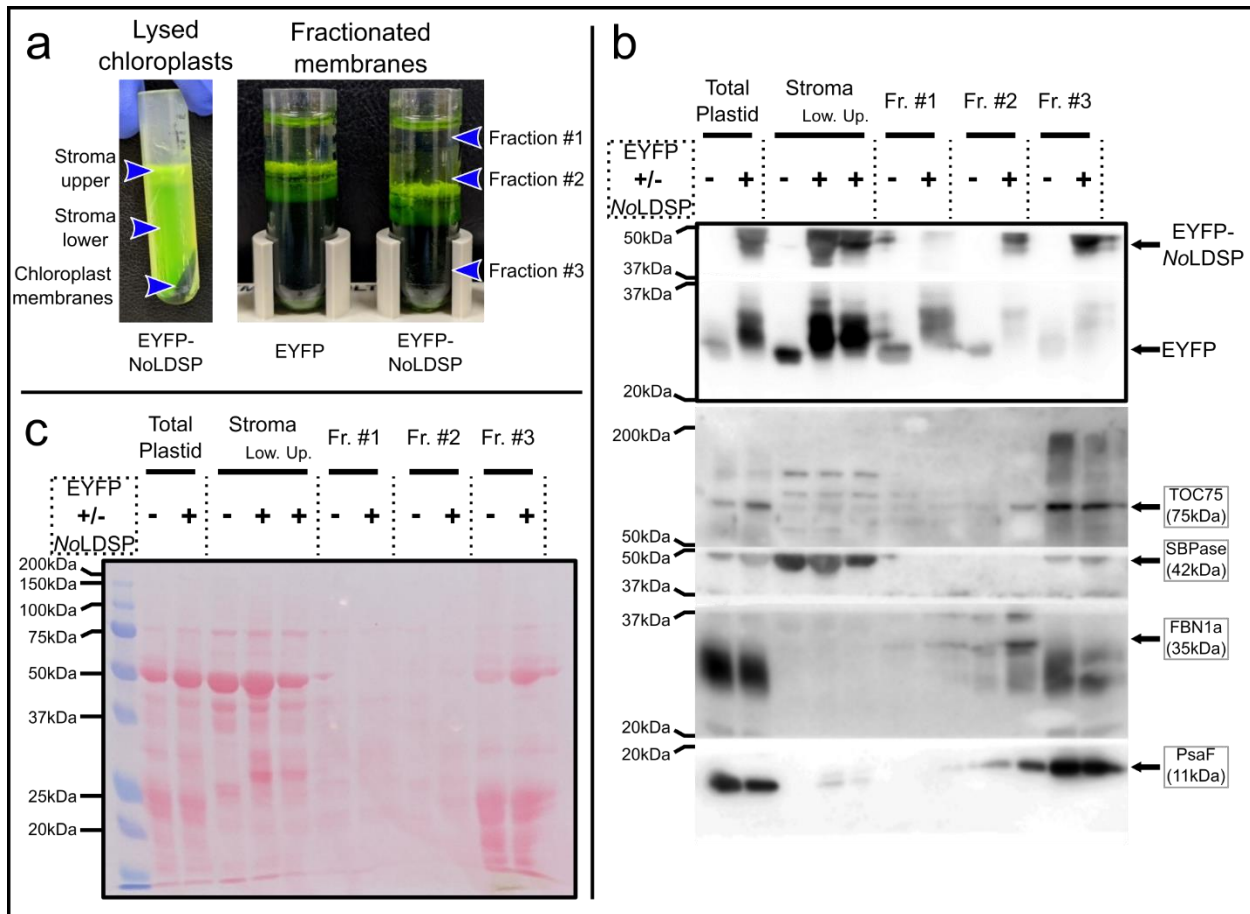


Figure A.3: Additional plastid fractionation and western blots demonstrating EYFP-NoLDSP localization in chloroplast membranes. During chloroplast isolation and fractionation (a), a possible upper layer in the stroma was included for analysis. The nitrocellulose membrane was cut after Ponceau S dye staining (c) to form fragments which could be visualized with each fraction marker by the indicated antibodies (b). The 20 – 37 kDa and 37 kDa – 50 kDa fragments in (b) were first visualized by the fraction specific antibody then washed and re-visualized with anti-GFP. Each lane is indicated as uninfiltreated, wild type plants (WT) or the presence of EYFP with (+) or without (-) fusion to NoLDSP.

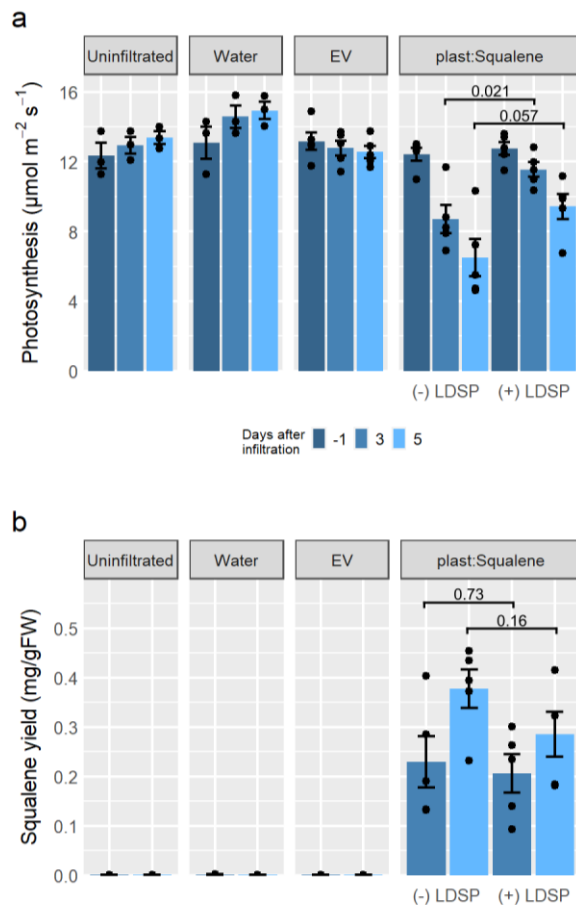


Figure A.4: Comparison of the effects of transient expression of plastid targeted squalene pathways in leaves and subsequent effects on photosynthesis compared to controls. Alongside plants expressing the plast:Squalene pathway with (+LDSP) and without (-LDSP) NoLDSP scaffolding, controls were included for uninfiltrated plants, plants infiltrated with water + 200 μM acetosyringone, and plants infiltrated with *Agrobacterium* harboring the pEAQ-HT empty vector (EV). Black circles show individual data points and bars represent means $\pm$ standard error. $n = 4$ plants per treatment. Individual t test statistical comparisons between means are shown by brackets and the indicated p -value.

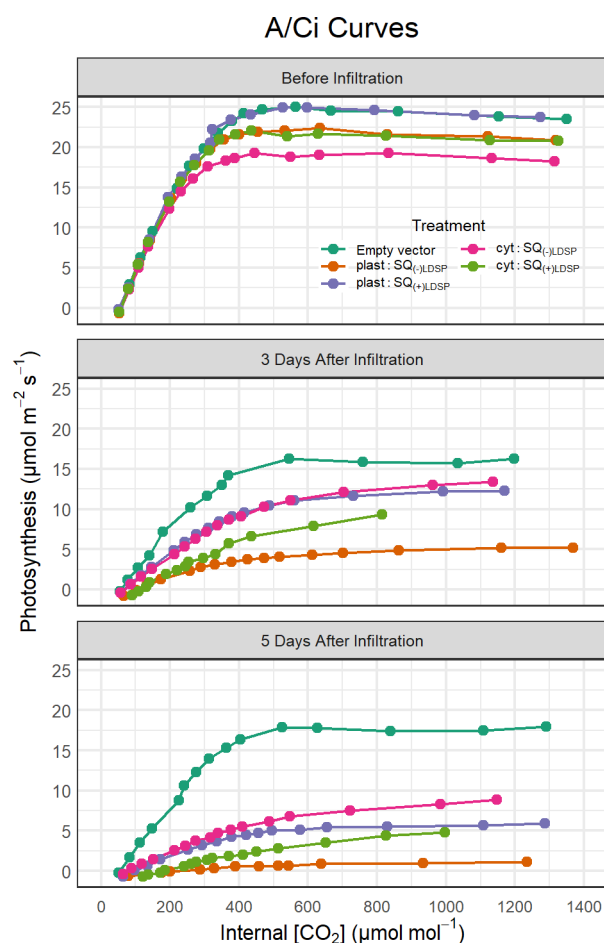


Figure A.5: A/Ci curves comparing plastid targeted and cytosolic squalene pathways, with and without NoLDSP scaffolding. Each curve was generated from a representative plant for each treatment indicated by color. Data points in each curve represent photosynthesis measured at the indicated internal CO_2 concentration under a saturating light intensity of $1000 \mu\text{mol m}^{-2} \text{s}^{-1}$. Maximum carboxylation rate (V_{cmax}), maximum rate of electron transport (J), and triose phosphate utilization rate (TPU) determined by fitting the Farquhar-von Caemmerer-Berry biochemical model of photosynthesis to A/Ci curves, are presented in Table A.2.

APPENDIX B

Supplemental Data for Chapter 3

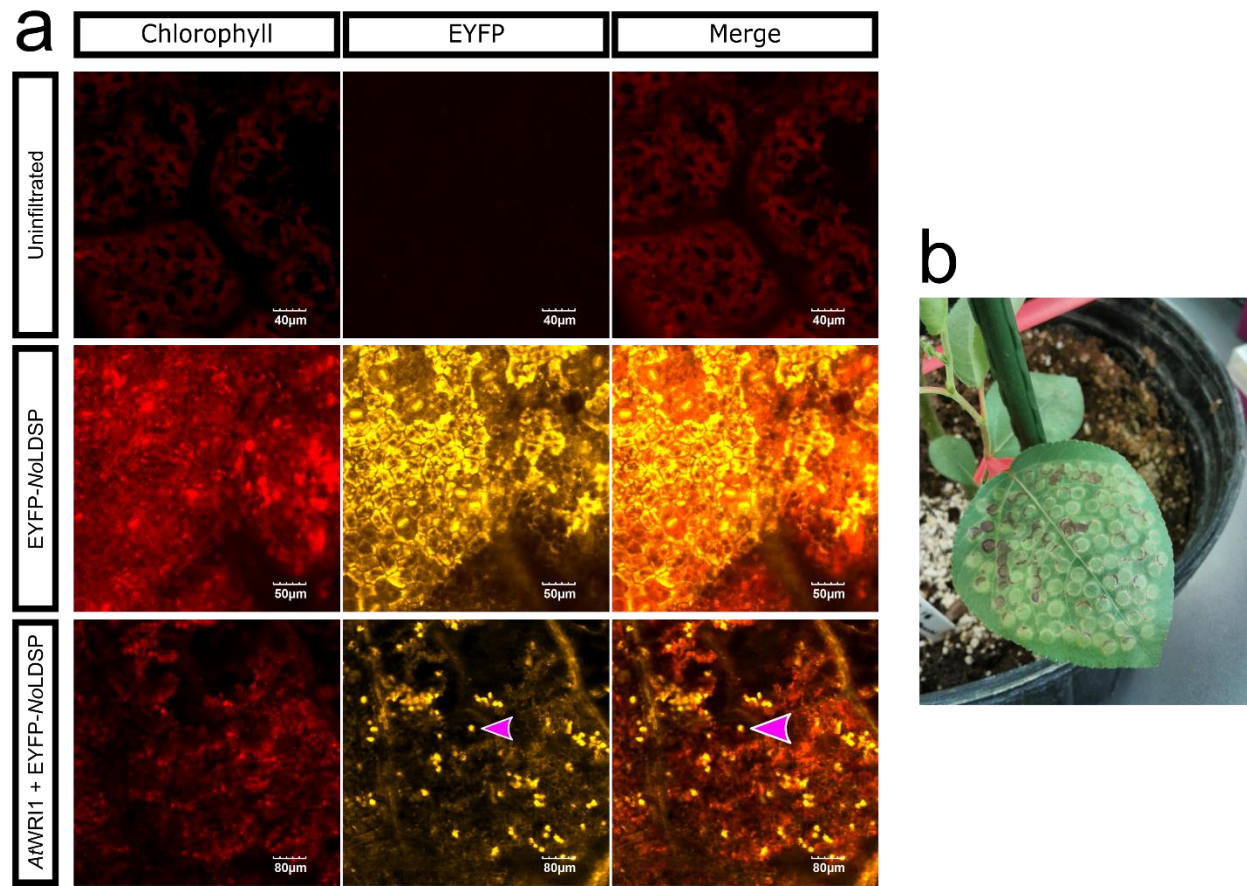


Figure B.1: Transient expression of lipid droplet scaffolding in poplar NM6 leaves. Panel (a) shows the fluorescent images of poplar NM6 plants either uninfiltrated, transiently expressing *EYFP-NoLDSP* fusions, or transiently expressing *AtWR11* and *EYFP-NoLDSP*. Panel (b) is a representative infiltrated poplar NM6 leaf showing the infiltrated areas around the syringe marks that were imaged. Pink arrows point to *EYFP-NoLDSP* anchored to lipid droplets.

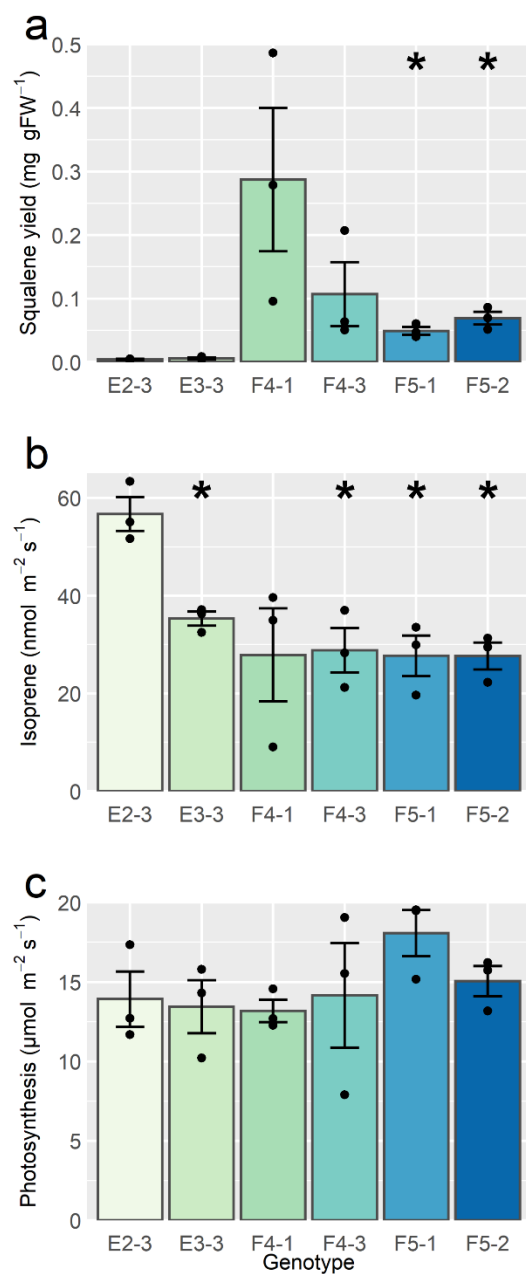
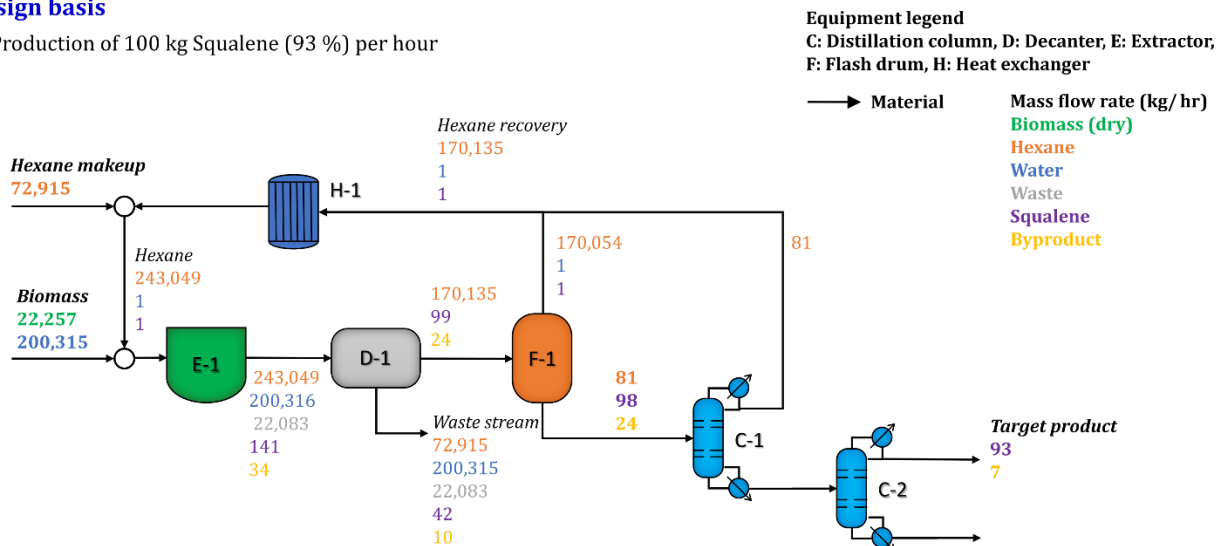


Figure B.2: Additional analysis of isoprene emission and photosynthesis in squalene producing poplar NM6. Squalene production (a), isoprene emission (b), and photosynthesis (c) measured for select transgenic poplar clones. For each plant the seventh fully formed leaf was selected from three branches to measure isoprene emission and photosynthesis with biological triplicates. Following measurements, the measured leaves were collected for squalene extraction. Each dot represents individual measurements and error bars represent the standard error. Asterisk indicates statistical significance compared to E2-3 ($p < 0.05$) as determined by t test.

a

Design basis

- Production of 100 kg Squalene (93 %) per hour



b

Simulation results

- Total required energy = 147 GJ/hr

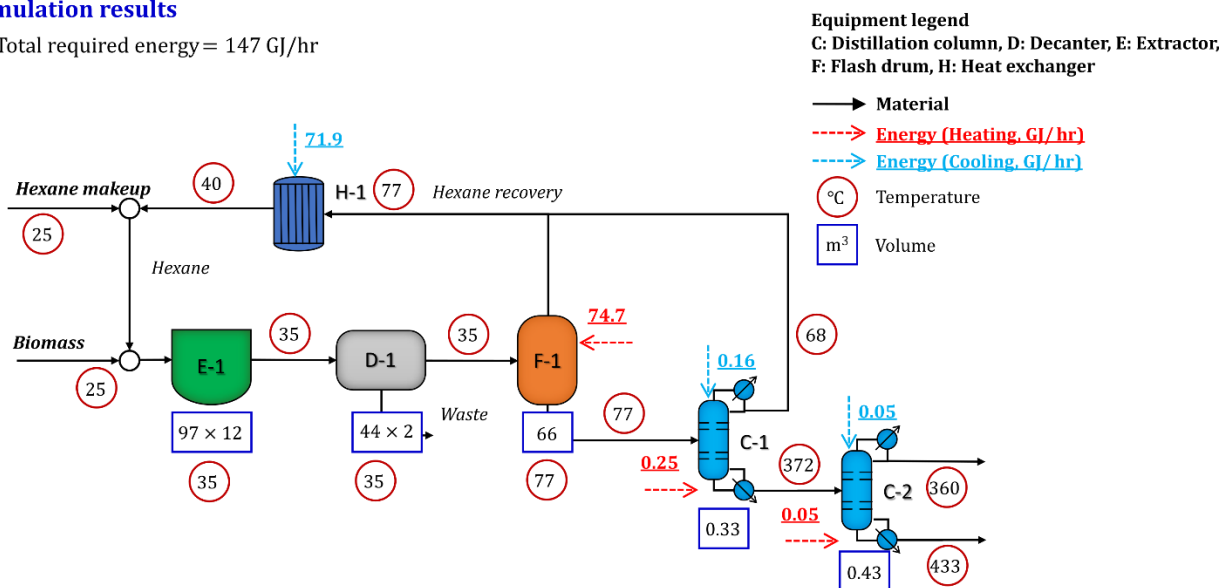


Figure B.3: Parameters used to the technoeconomic analyses performed for squalene extraction and purification from poplar leaves. The design basis (a) indicates the input and output values used for each step in the simulation of squalene extraction from bulk poplar leaves. In (b), the simulation of utility costs is determined for each step involved processing squalene. The following assumptions were used in the analysis represented: (i) 0.63 mg/gFW squalene yields, (ii) 90% leaf water content, (iii) 70% hexane recovery from tissue, (iv) hexane cost of \$0.89/kg, (v) feedstock cost of \$0.04/kg, and (vi) final squalene purity of 93%.

Table B.1: Economic parameters and assumptions used in technoeconomic analyses. Indicated are the assumptions included in the analyses and where they were derived from.

Biomass (\$ per ton)	40	For harvesting and collection ¹⁹⁵
Hexane (\$ per ton)	890	¹⁹⁶
Cooling water (\$ per GJ)	0.212	Aspen Energy Analyzer V11
Low pressure steam (\$ per GJ)	1.9	Aspen Energy Analyzer V11
High pressure steam (\$ per GJ)	2.5	Aspen Energy Analyzer V11
Plant operating hours per year	8410	¹⁹⁷
Plant life (year)	30	¹⁹⁷
Internal rate of return (%)	10	¹⁹⁷

APPENDIX C

Supplemental Data for Chapter 4

Table C.1: Promoters characterized in this study, the associated gene IDs, and reported native expression of each. Values are mean log2(1 + FPKM) of raw FPKM values reported across the different reported conditions for each tissue type.

Promoter	Putative BDP size	(-) Gene ID	(+) Gene ID	(-) Gene leaf FPKM	(+) Gene leaf FPKM	(-) Gene stem FPKM	(+) Gene stem FPKM	(-) Gene root FPKM	(+) Gene root FPKM
<i>Pt</i> BDP-1	319	Potri.001G136800	Potri.001G136900	2.43	2.74	2.26	2.88	2.23	1.56
<i>Pt</i> BDP-2	140	Potri.002G036400	Potri.002G036500	2.97	2.49	3.20	2.42	3.60	2.24
<i>Pt</i> BDP-3	254	Potri.004G188700	Potri.004G188800	4.69	3.81	4.38	3.61	4.33	4.35
<i>Pt</i> BDP-4	891	Potri.004G203700	Potri.004G203800	3.33	3.16	3.36	3.03	4.45	2.88
<i>Pt</i> BDP-5	964	Potri.005G179600	Potri.005G179700	2.69	2.92	2.83	2.56	2.60	1.96
<i>Pt</i> BDP-6	182	Potri.005G226300	Potri.005G226400	3.51	3.68	3.43	3.84	3.37	3.73
<i>Pt</i> BDP-7	489	Potri.006G000200	Potri.006G000300	3.20	3.13	3.27	3.32	3.38	3.80
<i>Pt</i> BDP-8	314	Potri.006G221900	Potri.006G222000	2.29	2.47	2.12	2.50	0.37	2.27
<i>Pt</i> BDP-9	978	Potri.006G250300	Potri.006G250400	3.33	3.04	3.46	3.24	3.37	3.34
<i>Pt</i> BDP-10	690	Potri.007G115800	Potri.007G115900	3.12	3.42	3.06	3.84	2.93	5.03
<i>Pt</i> BDP-11	556	Potri.008G003100	Potri.008G003200	2.76	3.28	2.73	3.10	2.71	3.51
<i>Pt</i> BDP-12	31	Potri.008G097300	Potri.008G097400	2.74	2.38	2.19	2.12	1.64	2.32
<i>Pt</i> BDP-13	278	Potri.009G165300	Potri.009G165400	3.38	3.01	3.18	2.56	3.30	2.30
<i>Pt</i> BDP-14	642	Potri.003G173800	Potri.003G173900	3.32	2.32	3.13	1.97	2.85	2.68
<i>Pt</i> BDP-15	934	Potri.004G004100	Potri.004G004200	3.38	2.21	3.28	2.37	3.25	2.60
<i>Pt</i> BDP-16	488	Potri.005G049400	Potri.005G049500	5.95	3.09	5.55	3.27	5.72	1.58
<i>Pt</i> BDP-17	420	Potri.009G125000	Potri.009G125100	6.05	2.93	5.93	2.65	5.63	2.61
<i>Pt</i> BDP-18	857	Potri.017G122500	Potri.017G122600	2.17	2.37	2.27	2.22	1.91	2.53
<i>Pt</i> BDP-19	905	Potri.013G020700	Potri.013G020800	2.22	2.19	1.63	1.87	1.82	0.64
<i>Pt</i> BDP-20	822	Potri.009G065800	Potri.009G065900	6.37	4.74	6.03	4.09	6.25	2.26
<i>Pt</i> BDP-21	298	Potri.002G100200	Potri.002G100300	5.10	3.49	4.46	3.30	4.12	2.74
<i>Pt</i> BDP-22	790	Potri.013G125500	Potri.013G125600	2.01	2.71	1.63	2.30	0.77	1.10
<i>Pt</i> BDP-23	1154	Potri.006G025200	Potri.006G025300	3.87	2.24	4.21	2.06	0.93	1.72
<i>Pt</i> BDP-24	876	Potri.005G239200	Potri.005G239300	6.53	11.20	6.00	10.45	1.33	3.44
<i>At</i> BDP-1	250	AT1G14270	AT1G14280	2.79	4.89	n/a	n/a	0.47	0.44
<i>At</i> BDP-2	1791	AT4G22505	AT4G22513	2.13	3.18	n/a	n/a	0.01	0.08
<i>At</i> BDP-3	1149	AT4G23130	AT4G23140	1.87	3.13	n/a	n/a	0.23	0.11
<i>At</i> BDP-4	689	AT4G23290	AT4G23300	2.45	3.51	n/a	n/a	0.02	0.01
<i>At</i> BDP-5	210	AT5G15850	AT5G15853	2.92	1.15	n/a	n/a	0.09	0.12
<i>At</i> BDP-6	1567	AT5G18020	AT5G18030	1.27	1.84	n/a	n/a	0.00	0.00
<i>At</i> BDP-7	693	AT5G18050	AT5G18060	1.18	1.66	n/a	n/a	0.00	0.00
<i>At</i> BDP-8	222	AT5G27290	AT5G27300	2.77	1.57	n/a	n/a	0.00	0.16
<i>At</i> BDP-9	221	AT5G35480	AT5G35490	5.79	4.48	n/a	n/a	0.00	0.03
<i>At</i> BDP-10	651	AT5G38510	AT5G38520	3.34	5.06	n/a	n/a	0.36	0.15
Promoter	Gene ID	Leaf FPKM	Stem FPKM	Root FPKM					
<i>Pt</i> UDP1	Potri.004G054600	4.28	1.64	0.84					
<i>Pt</i> UDP2	Potri.013G115900	4.40	1.89	1.00					
<i>Pt</i> UDP3	Potri.019G015300	4.55	1.10	0.96					

Table C.2: Primers used to amplify candidate promoters from genomic DNA. Red letters indicate where a point mutation was introduced in the primer to remove a start codon in the flanking gene if the primer needed to be designed in a portion of the gene.

Primer name	Primer sequence	Primer name	Primer sequence
PtBDP1-Fwd	TTTCAATCTCACAGTCCACACAC	AtBDP1-Fwd	TACACAGAGAATAGTCACTGAACATTAAC
PtBDP1-Rev	TTTTTATAATTAGTTTTTGGAACTGGAGGAG	AtBDP1-Rev	GGTACCAGTTTGTCTTTTGTGTAT
PtBDP2-Fwd	TTGTAGCCGCCGGAAGTG	AtBDP2-Fwd	GATTCAGTTTGTCCGAGGTATATATATATGTCCTT
PtBDP2-Rev	TGTTCCAGCTCCTTATTACCTTCTAC	AtBDP2-Rev	CAGATTGTGAAAATGCTAAATAAGCTTTTGATG
PtBDP3-Fwd	CTCTTTTCTCTTTCAAGATCTACTCC	AtBDP3-Fwd	CAGTATCCAATTTCTCACCTTTCTCTC
PtBDP3-Rev	GTTTAAATGAGAGTAGAGGGTTTCAAGGG	AtBDP3-Rev	TGTTTGAGTACCTTGATTGAGTTTCAC
PtBDP4-Fwd	TGCTGAAAATTGGGAAGGGTGC	AtBDP4-Fwd	CGAGTTTGATTTCTTGTTAAACCTCC
PtBDP4-Rev	TGATGATGAAAAAGAGGAGAGACGC	AtBDP4-Rev	TTTGCTTACGAGATTATAGCTGTGAT
PtBDP5-Fwd	GGTTCCCGAGACCAGCAGG	AtBDP5-Fwd	ACTAACCGCAGATGTTGAGTTGT
PtBDP5-Rev	TTCTGCTTGCTAACCCACCTGC	AtBDP5-Rev	TCGAAGATAACAAATGGTTAAGAGTCATAATC
PtBDP6-Fwd	TGTAGCAACTCTGAACAAACAAAGTC	AtBDP6-Fwd	CTCACAAAAGCCAGATCTATTGTTTTG
PtBDP6-Rev	TTTTGATGTTGTTGTTTGGTATTGACG	AtBDP6-Rev	ATCTGTTATTTTGTACGAAAAGGTTTTGAAG
PtBDP7-Fwd	CGTCCGGTCACTATTCTCTCTG	AtBDP7-Fwd	CAGAGCCAGTTATAATTCTCTATTGTCTG
PtBDP7-Rev1	CTTCTTTATTTTAATAATAGTAGCACCAATAATCTCTCCTC	AtBDP7-Rev	GCCAGTTCTTTTGTCTGAAAGTTTGA
PtBDP7-Rev2	CATACTCATGATCTACTCTATGAGCCA CTT	AtBDP8-Fwd	TCAGCAATAAAGGATTTGTGAACTAAATC
PtBDP8-Fwd1	GGTCTGTGACCGTTTTTTCTCTG	AtBDP8-Rev	CAGCCAGTTTACAATGAGTCACCTA
PtBDP8-Fwd2	CACAGATATGTCCTGTGACCG	AtBDP9-Fwd	TTCAGATTTTTAATTAGAAAAAGAGTTGGTTG
PtBDP8-Rev	CCCACTTTAATTTTCGCTTTTCTCTCAG	AtBDP9-Rev	TGCTGCATTGTCATCTACCAAC
PtBDP9-Fwd	ATAGAAGCTAGGAGGGTTTCAAATTTCA	AtBDP10-Fwd	CTTCAGGTCCTGTCATTTTCAGA
PtBDP9-Rev	TTGATTTTAAGATTAAATTCATTTTGAATTGGTAGA	AtBDP10-Rev	CGCTTTTTTCTCTCTTCTCTTTGATCT
PtBDP10-Fwd	TGAAGCTGAATCGAAACGTGTCTA		
PtBDP10-Rev	CTTGATGATATTTTCTTTGTGATTATGTGGCTG		
PtBDP11-Fwd1	CTTTTTTCAAAAAAGAAAAAATCAACCCAAACCC		
PtBDP11-Fwd2	CCCCTAAATTGATCTCCCA CTTTTTTC		
PtBDP11-Rev	CACTACTGACGAGGACGCTGAT		
PtBDP12-Fwd	GATGGTTCTGGGGTTTAAGAAAACGAG		
PtBDP12-Rev	ATCAGCAATGGATAACGGTGCTGG		
PtBDP13-Fwd1	AACAGTAGGAAAAAGAAAAAATCTGAACCTAGTG		
PtBDP13-Fwd2	GCAGATTGAGAGCCA AACAGTAGG		
PtBDP13-Rev	AGCCAGAATGCTCTCTCTCTCC		
PtBDP14-Fwd1	TTTCTCTTGTGTTGCTCTTCTCTCTCT		
PtBDP14-Fwd2	CCGTCACTTCTCTTGTGTTGCTC		
PtBDP14-Rev	AGTATCGCCAAGAAGATTACGAGAGAAG		
PtBDP15-Fwd	CACAAAAGCTTTGCTCAATAACGC		
PtBDP15-Rev	TGCTTTTACTTTAGATTGATTGCGGAGAC		
PtBDP16-Fwd	GGCTTCAATCAAAACCTAGATTTTGC		
PtBDP16-Rev	TGAAGACTTTGAAGTCTGGGGTGATG		
PtBDP17-Fwd	GGATATAAACGTCGCACGTGTTAAA		
PtBDP17-Rev	TTTTCTCCACAACAAACTCAGTCTATCC		
PtBDP18-Fwd	CTCTTTTCTTCTCTGATCTCTCAGTGAC		
PtBDP18-Rev	TCTGTTCTCTCTCTCTCTCTGCTTTG		
PtBDP19-Fwd	GTCTACCGTCAGATCTGAGAGAAATTGAG		
PtBDP19-Rev	TGAAGCTAAGCTAACTACCAACTGC		
PtBDP20-Fwd	TGATTCTGAATGAGGTTTCTCTCGC		
PtBDP20-Rev	ATCCGGAATTTCTCTCAAGTTACCATC		
PtBDP21-Fwd	CTCGAAAGACTTGCCAAAACCTC		
PtBDP21-Rev	GCTATCCTCTCTCTCTCTCTCT		
PtBDP22-Fwd	CTTTAAGACCCGTTCTCTATAGAATG		
PtBDP22-Rev	CAAAGCTTGCTTCTCTCTCTCT		
PtBDP23-Fwd	TTAGAACGTGGGGAGGAGGAGGA		
PtBDP23-Rev	GGCGCAGTTAGTTACATGTGATTTTCG		
PtBDP24-Fwd	TTTCGAGCTGATGAGTAAACGTGTATGT		
PtBDP24-Rev	TGCTGGATGAGCGGGTGCA		

REFERENCES

REFERENCES

- (1) Gurib-Fakim, A. Medicinal Plants: Traditions of Yesterday and Drugs of Tomorrow. *Mol. Aspects Med.* **2006**, 27 (1), 1–93.
<https://doi.org/10.1016/j.mam.2005.07.008>.
- (2) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *J. Nat. Prod.* **2016**, 79 (3), 629–661.
<https://doi.org/10.1021/acs.jnatprod.5b01055>.
- (3) Afendi, F. M.; Okada, T.; Yamazaki, M.; Hirai-Morita, A.; Nakamura, Y.; Nakamura, K.; Ikeda, S.; Takahashi, H.; Altaf-Ul-Amin, Md.; Darusman, L. K.; Saito, K.; Kanaya, S. KNApSACK Family Databases: Integrated Metabolite–Plant Species Databases for Multifaceted Plant Research. *Plant Cell Physiol.* **2012**, 53 (2), e1. <https://doi.org/10.1093/pcp/pcr165>.
- (4) Dixon, R. A.; Strack, D. Phytochemistry Meets Genome Analysis, and Beyond..... *Phytochemistry* **2003**, 62 (6), 815–816.
[https://doi.org/10.1016/S0031-9422\(02\)00712-4](https://doi.org/10.1016/S0031-9422(02)00712-4).
- (5) Yuan, L.; Grotewold, E. Metabolic Engineering to Enhance the Value of Plants as Green Factories. *Metab. Eng.* **2015**, 27, 83–91.
<https://doi.org/10.1016/j.ymben.2014.11.005>.
- (6) Huchelmann, A.; Boutry, M.; Hachez, C. Plant Glandular Trichomes: Natural Cell Factories of High Biotechnological Interest. *Plant Physiol.* **2017**, 175 (1), 6–22.
<https://doi.org/10.1104/pp.17.00727>.
- (7) Chassagne, F.; Cabanac, G.; Hubert, G.; David, B.; Marti, G. The Landscape of Natural Product Diversity and Their Pharmacological Relevance from a Focus on the Dictionary of Natural Products®. *Phytochem. Rev.* **2019**, 18 (3), 601–622.
<https://doi.org/10.1007/s11101-019-09606-2>.
- (8) Johnson, S. R.; Bhat, W. W.; Bibik, J.; Turmo, A.; Hamberger, B.; Consortium, E. M. G.; Hamberger, B.; Boachon, B.; Buell, C. R.; Crisovan, E.; Dudareva, N.; Garcia, N.; Godden, G.; Henry, L.; Kamileen, M. O.; Kates, H. R.; Kilgore, M. B.; Lichman, B. R.; Mavrodiev, E. V.; Newton, L.; Rodriguez-Lopez, C.; O'Connor, S. E.; Soltis, D.; Soltis, P.; Vaillancourt, B.; Wiegert-Rininger, K.; Zhao, D. A Database-Driven Approach Identifies Additional Diterpene Synthase Activities in the Mint Family (Lamiaceae). *J. Biol. Chem.* **2019**, 294 (4), 1349–1362.
<https://doi.org/10.1074/jbc.RA118.006025>.
- (9) Karunanithi, P. S.; Zerbe, P. Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Front. Plant Sci.* **2019**, 10.
<https://doi.org/10.3389/fpls.2019.01166>.

- (10) Miller, G. P.; Bhat, W. W.; Lanier, E. R.; Johnson, S. R.; Mathieu, D. T.; Hamberger, B. The Biosynthesis of the Anti-Microbial Diterpenoid Leubethanol in *Leucophyllum Frutescens* Proceeds via an All-Cis Prenyl Intermediate. *Plant J.* **2020**, *104* (3), 693–705. <https://doi.org/10.1111/tpj.14957>.
- (11) Pichersky, E.; Raguso, R. A. Why Do Plants Produce so Many Terpenoid Compounds? *New Phytol.* **2018**, *220* (3), 692–702. <https://doi.org/10.1111/nph.14178>.
- (12) Wu, S.; Schalk, M.; Clark, A.; Miles, R. B.; Coates, R.; Chappell, J. Redirection of Cytosolic or Plastidic Isoprenoid Precursors Elevates Terpene Production in Plants. *Nat. Biotechnol.* **2006**, *24* (11), 1441–1447. <https://doi.org/10.1038/nbt1251>.
- (13) Wu, S.; Jiang, Z.; Kempinski, C.; Eric Nybo, S.; Husodo, S.; Williams, R.; Chappell, J. Engineering Triterpene Metabolism in Tobacco. *Planta* **2012**, *236* (3), 867–877. <https://doi.org/10.1007/s00425-012-1680-4>.
- (14) Zhao, C.; Kim, Y.; Zeng, Y.; Li, M.; Wang, X.; Hu, C.; Gorman, C.; Dai, S. Y.; Ding, S.-Y.; Yuan, J. S. Co-Compartmentation of Terpene Biosynthesis and Storage via Synthetic Droplet. *ACS Synth. Biol.* **2018**, *7* (3), 774–781. <https://doi.org/10.1021/acssynbio.7b00368>.
- (15) Sadre, R.; Kuo, P.; Chen, J.; Yang, Y.; Banerjee, A.; Benning, C.; Hamberger, B. Cytosolic Lipid Droplets as Engineered Organelles for Production and Accumulation of Terpenoid Biomaterials in Leaves. *Nat. Commun.* **2019**, *10* (1), 853. <https://doi.org/10.1038/s41467-019-08515-4>.
- (16) Zerbe, P.; Bohlmann, J. Plant Diterpene Synthases: Exploring Modularity and Metabolic Diversity for Bioengineering. *Trends Biotechnol.* **2015**, *33* (7), 419–428. <https://doi.org/10.1016/j.tibtech.2015.04.006>.
- (17) Andersen-Ranberg, J.; Kongstad, K. T.; Nielsen, M. T.; Jensen, N. B.; Pateraki, I.; Bach, S. S.; Hamberger, B.; Zerbe, P.; Staerk, D.; Bohlmann, J.; Møller, B. L.; Hamberger, B. Expanding the Landscape of Diterpene Structural Diversity through Stereochemically Controlled Combinatorial Biosynthesis. *Angew. Chem. Int. Ed Engl.* **2016**, *55* (6), 2142–2146. <https://doi.org/10.1002/anie.201510650>.
- (18) Yang, S.; Tian, H.; Sun, B.; Liu, Y.; Hao, Y.; Lv, Y. One-Pot Synthesis of (–)-Ambrox. *Sci. Rep.* **2016**, *6*, 32650. <https://doi.org/10.1038/srep32650>.
- (19) Ke, D.; Caiyin, Q.; Zhao, F.; Liu, T.; Lu, W. Heterologous Biosynthesis of Triterpenoid Ambrein in Engineered *Escherichia Coli*. *Biotechnol. Lett.* **2018**, *40* (2), 399–404. <https://doi.org/10.1007/s10529-017-2483-2>.
- (20) Moser, S.; Strohmeier, G. A.; Leitner, E.; Plocek, T. J.; Vanhessche, K.; Pichler, H. Whole-Cell (+)-Ambrein Production in the Yeast *Pichia Pastoris*. *Metab. Eng. Commun.* **2018**, *7*, e00077. <https://doi.org/10.1016/j.mec.2018.e00077>.

- (21) Croteau, R.; Ketchum, R. E. B.; Long, R. M.; Kaspera, R.; Wildung, M. R. Taxol Biosynthesis and Molecular Genetics. *Phytochem. Rev. Proc. Phytochem. Soc. Eur.* **2006**, *5* (1), 75–97. <https://doi.org/10.1007/s11101-005-3748-2>.
- (22) Xu, J.; Dolan, M. C.; Medrano, G.; Cramer, C. L.; Weathers, P. J. Green Factory: Plants as Bioproduction Platforms for Recombinant Proteins. *Biotechnol. Adv.* **2012**, *30* (5), 1171–1184. <https://doi.org/10.1016/j.biotechadv.2011.08.020>.
- (23) Tiwari, P.; Khare, T.; Shriram, V.; Bae, H.; Kumar, V. Plant Synthetic Biology for Producing Potent Phyto-Antimicrobials to Combat Antimicrobial Resistance. *Biotechnol. Adv.* **2021**, *48*, 107729. <https://doi.org/10.1016/j.biotechadv.2021.107729>.
- (24) Busch, M.; Seuter, A.; Hain, R. Functional Analysis of the Early Steps of Carotenoid Biosynthesis in Tobacco. *Plant Physiol.* **2002**, *128* (2), 439–453. <https://doi.org/10.1104/pp.010573>.
- (25) Vickers, C. E.; Possell, M.; Laothawornkitkul, J.; Ryan, A. C.; Hewitt, C. N.; Mullineaux, P. M. Isoprene Synthesis in Plants: Lessons from a Transgenic Tobacco Model. *Plant Cell Environ.* **2011**, *34* (6), 1043–1053. <https://doi.org/10.1111/j.1365-3040.2011.02303.x>.
- (26) Hasan, Md. M.; Kim, H.-S.; Jeon, J.-H.; Kim, S. H.; Moon, B.; Song, J.-Y.; Shim, S. H.; Baek, K.-H. Metabolic Engineering of *Nicotiana Benthiana* for the Increased Production of Taxadiene. *Plant Cell Rep.* **2014**, *33* (6), 895–904. <https://doi.org/10.1007/s00299-014-1568-9>.
- (27) Reed, J.; Stephenson, M. J.; Miettinen, K.; Brouwer, B.; Leveau, A.; Brett, P.; Goss, R. J. M.; Goossens, A.; O'Connell, M. A.; Osbourn, A. A Translational Synthetic Biology Platform for Rapid Access to Gram-Scale Quantities of Novel Drug-like Molecules. *Metab. Eng.* **2017**, *42*, 185–193. <https://doi.org/10.1016/j.ymben.2017.06.012>.
- (28) Andersen, T. B.; Llorente, B.; Morelli, L.; Torres-Montilla, S.; Bordanaba-Florit, G.; Espinosa, F. A.; Rodriguez-Goberna, M. R.; Campos, N.; Olmedilla-Alonso, B.; Llansola-Portoles, M. J.; Pascal, A. A.; Rodriguez-Concepcion, M. An Engineered Extrplastidial Pathway for Carotenoid Biofortification of Leaves. *Plant Biotechnol. J.* **2021**, *19* (5), 1008–1021. <https://doi.org/10.1111/pbi.13526>.
- (29) Kovacs, K.; Zhang, L.; Linforth, R. S. T.; Whittaker, B.; Hayes, C. J.; Fray, R. G. Redirection of Carotenoid Metabolism for the Efficient Production of Taxadiene [Taxa-4(5),11(12)-Diene] in Transgenic Tomato Fruit. *Transgenic Res.* **2007**, *16* (1), 121–126. <https://doi.org/10.1007/s11248-006-9039-x>.
- (30) Henry, L. K.; Gutensohn, M.; Thomas, S. T.; Noel, J. P.; Dudareva, N. Orthologs of the Archaeal Isopentenyl Phosphate Kinase Regulate Terpenoid Production in Plants. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (32), 10050–10055. <https://doi.org/10.1073/pnas.1504798112>.

- (31) Henry, L. K.; Thomas, S. T.; Widhalm, J. R.; Lynch, J. H.; Davis, T. C.; Kessler, S. A.; Bohlmann, J.; Noel, J. P.; Dudareva, N. Contribution of Isopentenyl Phosphate to Plant Terpenoid Metabolism. *Nat. Plants* **2018**, *4* (9), 721. <https://doi.org/10.1038/s41477-018-0220-z>.
- (32) Zhan, X.; Zhang, Y.-H.; Chen, D.-F.; Simonsen, H. T. Metabolic Engineering of the Moss *Physcomitrella Patens* to Produce the Sesquiterpenoids Patchoulol and α/β -Santalene. *Front. Plant Sci.* **2014**, *5*.
- (33) Banerjee, A.; Arnesen, J. A.; Moser, D.; Motsa, B. B.; Johnson, S. R.; Hamberger, B. Engineering Modular Diterpene Biosynthetic Pathways in *Physcomitrella Patens*. *Planta* **2019**, *249* (1), 221–233. <https://doi.org/10.1007/s00425-018-3053-0>.
- (34) Harker, M.; Holmberg, N.; Clayton, J. C.; Gibbard, C. L.; Wallace, A. D.; Rawlins, S.; Hellyer, S. A.; Lanot, A.; Safford, R. Enhancement of Seed Phytosterol Levels by Expression of an N-Terminal Truncated *Hevea Brasiliensis* (Rubber Tree) 3-Hydroxy-3-Methylglutaryl-CoA Reductase. *Plant Biotechnol. J.* **2003**, *1* (2), 113–121. <https://doi.org/10.1046/j.1467-7652.2003.00011.x>.
- (35) Banerjee, A.; Preiser, A. L.; Sharkey, T. D. Engineering of Recombinant Poplar Deoxy-D-Xylulose-5-Phosphate Synthase (PtDXS) by Site-Directed Mutagenesis Improves Its Activity. *PLOS ONE* **2016**, *11* (8), e0161534. <https://doi.org/10.1371/journal.pone.0161534>.
- (36) Lee, A.-R.; Kwon, M.; Kang, M.-K.; Kim, J.; Kim, S.-U.; Ro, D.-K. Increased Sesqui- and Triterpene Production by Co-Expression of HMG-CoA Reductase and Biotin Carboxyl Carrier Protein in Tobacco (*Nicotiana Benthiana*). *Metab. Eng.* **2019**, *52*, 20–28. <https://doi.org/10.1016/j.ymben.2018.10.008>.
- (37) Wright, L. P.; Rohwer, J. M.; Ghirardo, A.; Hammerbacher, A.; Ortiz-Alcaide, M.; Raguschke, B.; Schnitzler, J.-P.; Gershenzon, J.; Phillips, M. A. Deoxyxylulose 5-Phosphate Synthase Controls Flux through the Methylerythritol 4-Phosphate Pathway in *Arabidopsis*. *Plant Physiol.* **2014**, *165* (4), 1488–1504. <https://doi.org/10.1104/pp.114.245191>.
- (38) Chappell, J.; Wolf, F.; Proulx, J.; Cuellar, R.; Saunders, C. Is the Reaction Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase a Rate-Limiting Step for Isoprenoid Biosynthesis in Plants? *Plant Physiol.* **1995**, *109* (4), 1337–1343.
- (39) Gutensohn, M.; Henry, L. K.; Gentry, S. A.; Lynch, J. H.; Nguyen, T. T. H.; Pichersky, E.; Dudareva, N. Overcoming Bottlenecks for Metabolic Engineering of Sesquiterpene Production in Tomato Fruits. *Front. Plant Sci.* **2021**, *12*, 691754. <https://doi.org/10.3389/fpls.2021.691754>.
- (40) Kirby, J.; Nishimoto, M.; Chow, R. W. N.; Baidoo, E. E. K.; Wang, G.; Martin, J.; Schackwitz, W.; Chan, R.; Fortman, J. L.; Keasling, J. D. Enhancing Terpene

- Yield from Sugars via Novel Routes to 1-Deoxy-d-Xylulose 5-Phosphate. *Appl. Environ. Microbiol.* **2015**, *81* (1), 130–138. <https://doi.org/10.1128/AEM.02920-14>.
- (41) Leonard, E.; Ajikumar, P. K.; Thayer, K.; Xiao, W.-H.; Mo, J. D.; Tidor, B.; Stephanopoulos, G.; Prather, K. L. J. Combining Metabolic and Protein Engineering of a Terpenoid Biosynthetic Pathway for Overproduction and Selectivity Control. *Proc. Natl. Acad. Sci.* **2010**, *107* (31), 13654–13659. <https://doi.org/10.1073/pnas.1006138107>.
 - (42) Malhotra, K.; Subramaniyan, M.; Rawat, K.; Kalamuddin, Md.; Qureshi, M. I.; Malhotra, P.; Mohammed, A.; Cornish, K.; Daniell, H.; Kumar, S. Compartmentalized Metabolic Engineering for Artemisinin Biosynthesis and Effective Malaria Treatment by Oral Delivery of Plant Cells. *Mol. Plant* **2016**, *9* (11), 1464–1477. <https://doi.org/10.1016/j.molp.2016.09.013>.
 - (43) Stålberg, K.; Ellerstöm, M.; Ezcurra, I.; Ablov, S.; Rask, L. Disruption of an Overlapping E-Box/ABRE Motif Abolished High Transcription of the NapA Storage-Protein Promoter in Transgenic *Brassica Napus* Seeds. *Planta* **1996**, *199* (4). <https://doi.org/10.1007/BF00195181>.
 - (44) Li, Y.; Liu, S.; Yu, Z.; Liu, Y.; Wu, P. Isolation and Characterization of Two Novel Root-Specific Promoters in Rice (*Oryza Sativa* L.). *Plant Sci.* **2013**, *207*, 37–44. <https://doi.org/10.1016/j.plantsci.2013.02.002>.
 - (45) Wang, R.; Yan, Y.; Zhu, M.; Yang, M.; Zhou, F.; Chen, H.; Lin, Y. Isolation and Functional Characterization of Bidirectional Promoters in Rice. *Front. Plant Sci.* **2016**, *7*. <https://doi.org/10.3389/fpls.2016.00766>.
 - (46) Liu, X.; Li, S.; Yang, W.; Mu, B.; Jiao, Y.; Zhou, X.; Zhang, C.; Fan, Y.; Chen, R. Synthesis of Seed-Specific Bidirectional Promoters for Metabolic Engineering of Anthocyanin-Rich Maize. *Plant Cell Physiol.* **2018**, *59* (10), 1942–1955. <https://doi.org/10.1093/pcp/pcy110>.
 - (47) Gao, L.; Tian, Y.; Chen, M.-C.; Wei, L.; Gao, T.-G.; Yin, H.-J.; Zhang, J.-L.; Kumar, T.; Liu, L.-B.; Wang, S.-M. Cloning and Functional Characterization of Epidermis-Specific Promoter MtML1 from *Medicago Truncatula*. *J. Biotechnol.* **2019**, *300*, 32–39. <https://doi.org/10.1016/j.jbiotec.2019.05.003>.
 - (48) Li, Y.; Dong, C.; Hu, M.; Bai, Z.; Tong, C.; Zuo, R.; Liu, Y.; Cheng, X.; Cheng, M.; Huang, J.; Liu, S. Identification of Flower-Specific Promoters through Comparative Transcriptome Analysis in *Brassica Napus*. *Int. J. Mol. Sci.* **2019**, *20* (23), 5949. <https://doi.org/10.3390/ijms20235949>.
 - (49) Bai, J.; Wang, X.; Wu, H.; Ling, F.; Zhao, Y.; Lin, Y.; Wang, R. Comprehensive Construction Strategy of Bidirectional Green Tissue-Specific Synthetic Promoters. *Plant Biotechnol. J.* **2020**, *18* (3), 668–678. <https://doi.org/10.1111/pbi.13231>.

- (50) Herpen, T. W. J. M. van; Cankar, K.; Nogueira, M.; Bosch, D.; Bouwmeester, H. J.; Beekwilder, J. *Nicotiana Benthamiana* as a Production Platform for Artemisinin Precursors. *PLOS ONE* **2010**, 5 (12), e14222. <https://doi.org/10.1371/journal.pone.0014222>.
- (51) Liu, Q.; Majdi, M.; Cankar, K.; Goedbloed, M.; Charnikhova, T.; Verstappen, F. W. A.; Vos, R. C. H. de; Beekwilder, J.; Krol, S. van der; Bouwmeester, H. J. Reconstitution of the Costunolide Biosynthetic Pathway in Yeast and *Nicotiana Benthamiana*. *PLOS ONE* **2011**, 6 (8), e23255. <https://doi.org/10.1371/journal.pone.0023255>.
- (52) Eljounaidi, K.; Cankar, K.; Comino, C.; Moglia, A.; Hehn, A.; Bourgaud, F.; Bouwmeester, H.; Menin, B.; Lanteri, S.; Beekwilder, J. Cytochrome P450s from *Cynara Cardunculus* L. CYP71AV9 and CYP71BL5, Catalyze Distinct Hydroxylations in the Sesquiterpene Lactone Biosynthetic Pathway. *Plant Sci.* **2014**, 223, 59–68. <https://doi.org/10.1016/j.plantsci.2014.03.007>.
- (53) Pateraki, I.; Andersen-Ranberg, J.; Hamberger, B.; Heskes, A. M.; Martens, H. J.; Zerbe, P.; Bach, S. S.; Møller, B. L.; Bohlmann, J.; Hamberger, B. Manoyl Oxide (13R), the Biosynthetic Precursor of Forskolin, Is Synthesized in Specialized Root Cork Cells in *Coleus Forskohlii*. *Plant Physiol.* **2014**, 164 (3), 1222–1236. <https://doi.org/10.1104/pp.113.228429>.
- (54) Boachon, B.; Junker, R. R.; Miesch, L.; Bassard, J.-E.; Höfer, R.; Caillieudeaux, R.; Seidel, D. E.; Lesot, A.; Heinrich, C.; Ginglinger, J.-F.; Allouche, L.; Vincent, B.; Wahyuni, D. S. C.; Paetz, C.; Beran, F.; Miesch, M.; Schneider, B.; Leiss, K.; Werck-Reichhart, D. CYP76C1 (Cytochrome P450)-Mediated Linalool Metabolism and the Formation of Volatile and Soluble Linalool Oxides in *Arabidopsis* Flowers: A Strategy for Defense against Floral Antagonists. *Plant Cell* **2015**, 27 (10), 2972–2990. <https://doi.org/10.1105/tpc.15.00399>.
- (55) Andersen, T. B.; Martinez-Swatson, K. A.; Rasmussen, S. A.; Boughton, B. A.; Jørgensen, K.; Andersen-Ranberg, J.; Nyberg, N.; Christensen, S. B.; Simonsen, H. T. Localization and In-Vivo Characterization of *Thapsia Garganica* CYP76AE2 Indicates a Role in Thapsigargin Biosynthesis. *Plant Physiol.* **2017**, 174 (1), 56–72. <https://doi.org/10.1104/pp.16.00055>.
- (56) Boachon, B.; Lynch, J. H.; Ray, S.; Yuan, J.; Caldo, K. M. P.; Junker, R. R.; Kessler, S. A.; Morgan, J. A.; Dudareva, N. Natural Fumigation as a Mechanism for Volatile Transport between Flower Organs. *Nat. Chem. Biol.* **2019**, 15 (6), 583. <https://doi.org/10.1038/s41589-019-0287-5>.
- (57) Muchlinski, A.; Chen, X.; Lovell, J. T.; Köllner, T. G.; Pelot, K. A.; Zerbe, P.; Ruggiero, M.; Callaway, L. I.; Laliberte, S.; Chen, F.; Tholl, D. Biosynthesis and Emission of Stress-Induced Volatile Terpenes in Roots and Leaves of Switchgrass (*Panicum Virgatum* L.). *Front. Plant Sci.* **2019**, 10. <https://doi.org/10.3389/fpls.2019.01144>.

- (58) Abbott, E.; Hall, D.; Hamberger, B.; Bohlmann, J. Laser Microdissection of Conifer Stem Tissues: Isolation and Analysis of High Quality RNA, Terpene Synthase Enzyme Activity and Terpenoid Metabolites from Resin Ducts and Cambial Zone Tissue of White Spruce (*Picea Glauca*). *BMC Plant Biol.* **2010**, *10* (1), 1–16. <https://doi.org/10.1186/1471-2229-10-106>.
- (59) Celedon, J. M.; Bohlmann, J. Oleoresin Defenses in Conifers: Chemical Diversity, Terpene Synthases and Limitations of Oleoresin Defense under Climate Change. *New Phytol.* **2019**, *224* (4), 1444–1463. <https://doi.org/10.1111/nph.15984>.
- (60) Kortbeek, R. W. J.; Xu, J.; Ramirez, A.; Spyropoulou, E.; Diergaarde, P.; Otten-Bruggeman, I.; de Both, M.; Nagel, R.; Schmidt, A.; Schuurink, R. C.; Bleeker, P. M. Chapter Twelve - Engineering of Tomato Glandular Trichomes for the Production of Specialized Metabolites. In *Methods in Enzymology*; O'Connor, S. E., Ed.; Synthetic Biology and Metabolic Engineering in Plants and Microbes Part B: Metabolism in Plants; Academic Press, 2016; Vol. 576, pp 305–331. <https://doi.org/10.1016/bs.mie.2016.02.014>.
- (61) Spokevicius, A. V.; Van Beveren, K.; Leitch, M. A.; Bossinger, G. *Agrobacterium*-Mediated in Vitro Transformation of Wood-Producing Stem Segments in Eucalypts. *Plant Cell Rep.* **2005**, *23* (9), 617–624. <https://doi.org/10.1007/s00299-004-0856-1>.
- (62) Zhong, L.; Zhang, Y.; Liu, H.; Sun, G.; Chen, R.; Song, S. *Agrobacterium*-Mediated Transient Expression via Root Absorption in Flowering Chinese Cabbage. *SpringerPlus* **2016**, *5* (1), 1825. <https://doi.org/10.1186/s40064-016-3518-1>.
- (63) Yang, M.; Baral, N. R.; Simmons, B. A.; Mortimer, J. C.; Shih, P. M.; Scown, C. D. Accumulation of High-Value Bioproducts in Planta Can Improve the Economics of Advanced Biofuels. *Proc. Natl. Acad. Sci.* **2020**, *117* (15), 8639–8648. <https://doi.org/10.1073/pnas.2000053117>.
- (64) Sannigrahi, P.; Ragauskas, A. J.; Tuskan, G. A. Poplar as a Feedstock for Biofuels: A Review of Compositional Characteristics. *Biofuels Bioprod. Biorefining* **2010**, *4* (2), 209–226. <https://doi.org/10.1002/bbb.206>.
- (65) Mathur, S.; Umakanth, A. V.; Tonapi, V. A.; Sharma, R.; Sharma, M. K. Sweet Sorghum as Biofuel Feedstock: Recent Advances and Available Resources. *Biotechnol. Biofuels* **2017**, *10* (1), 1–19. <https://doi.org/10.1186/s13068-017-0834-9>.
- (66) Costa, M. A.; Marques, J. V.; Dalisay, D. S.; Herman, B.; Bedgar, D. L.; Davin, L. B.; Lewis, N. G. Transgenic Hybrid Poplar for Sustainable and Scalable Production of the Commodity/Specialty Chemical, 2-Phenylethanol. *PLOS ONE* **2013**, *8* (12), e83169. <https://doi.org/10.1371/journal.pone.0083169>.

- (67) Lu, D.; Yuan, X.; Kim, S.-J.; Marques, J. V.; Chakravarthy, P. P.; Moinuddin, S. G. A.; Luchterhand, R.; Herman, B.; Davin, L. B.; Lewis, N. G. Eugenol Specialty Chemical Production in Transgenic Poplar (*Populus Tremula* × *P. Alba*) Field Trials. *Plant Biotechnol. J.* **2017**, *15* (8), 970–981. <https://doi.org/10.1111/pbi.12692>.
- (68) Schnitzler, J.-P.; Louis, S.; Behnke, K.; Loivamäki, M. Poplar Volatiles – Biosynthesis, Regulation and (Eco)Physiology of Isoprene and Stress-Induced Isoprenoids. *Plant Biol.* **2010**, *12* (2), 302–316. <https://doi.org/10.1111/j.1438-8677.2009.00284.x>.
- (69) Guenther, A. B.; Jiang, X.; Heald, C. L.; Sakulyanontvittaya, T.; Duhl, T.; Emmons, L. K.; Wang, X. The Model of Emissions of Gases and Aerosols from Nature Version 2.1 (MEGAN2.1): An Extended and Updated Framework for Modeling Biogenic Emissions. *Geosci. Model Dev.* **2012**, *5* (6), 1471–1492. <https://doi.org/10.5194/gmd-5-1471-2012>.
- (70) Vanzo, E.; Jud, W.; Li, Z.; Albert, A.; Domagalska, M. A.; Ghirardo, A.; Niederbacher, B.; Frenzel, J.; Beemster, G. T. S.; Asard, H.; Rennenberg, H.; Sharkey, T. D.; Hansel, A.; Schnitzler, J.-P. Facing the Future: Effects of Short-Term Climate Extremes on Isoprene-Emitting and Nonemitting Poplar. *Plant Physiol.* **2015**, *169* (1), 560–575. <https://doi.org/10.1104/pp.15.00871>.
- (71) Amack, S. C.; Antunes, M. S. CaMV35S Promoter – A Plant Biology and Biotechnology Workhorse in the Era of Synthetic Biology. *Curr. Plant Biol.* **2020**, *24*, 100179. <https://doi.org/10.1016/j.cpb.2020.100179>.
- (72) Peyret, H.; Lomonossoff, G. P. The PEAQ Vector Series: The Easy and Quick Way to Produce Recombinant Proteins in Plants. *Plant Mol. Biol.* **2013**, *83* (1–2), 51–58. <https://doi.org/10.1007/s11103-013-0036-1>.
- (73) Siddiqui, S. A.; Sarmiento, C.; Truve, E.; Lehto, H.; Lehto, K. Phenotypes and Functional Effects Caused by Various Viral RNA Silencing Suppressors in Transgenic *Nicotiana Benthamiana* and *N. Tabacum*. *Mol. Plant-Microbe Interactions®* **2008**, *21* (2), 178–187. <https://doi.org/10.1094/MPMI-21-2-0178>.
- (74) Halpin, C. Gene Stacking in Transgenic Plants – the Challenge for 21st Century Plant Biotechnology. *Plant Biotechnol. J.* **2005**, *3* (2), 141–155. <https://doi.org/10.1111/j.1467-7652.2004.00113.x>.
- (75) Bock, R. Strategies for Metabolic Pathway Engineering with Multiple Transgenes. *Plant Mol. Biol.* **2013**, *83* (1), 21–31. <https://doi.org/10.1007/s11103-013-0045-0>.
- (76) Collier, R.; Thomson, J. G.; Thilmony, R. A Versatile and Robust *Agrobacterium*-Based Gene Stacking System Generates High-Quality Transgenic *Arabidopsis* Plants. *Plant J. Cell Mol. Biol.* **2018**. <https://doi.org/10.1111/tpj.13992>.

- (77) Hirt, H.; Kögl, M.; Murbacher, T.; Heberle-Bors, E. Evolutionary Conservation of Transcriptional Machinery between Yeast and Plants as Shown by the Efficient Expression from the CaMV 35S Promoter and 35S Terminator. *Curr. Genet.* **1990**, *17* (6), 473–479. <https://doi.org/10.1007/BF00313074>.
- (78) Koncz, C.; De Greve, H.; André, D.; Deboeck, F.; Van Montagu, M.; Schell, J. The Opine Synthase Genes Carried by Ti Plasmids Contain All Signals Necessary for Expression in Plants. *EMBO J.* **1983**, *2* (9), 1597–1603.
- (79) Christensen, A. H.; Sharrock, R. A.; Quail, P. H. Maize Polyubiquitin Genes: Structure, Thermal Perturbation of Expression and Transcript Splicing, and Promoter Activity Following Transfer to Protoplasts by Electroporation. *Plant Mol. Biol.* **1992**, *18* (4), 675–689. <https://doi.org/10.1007/BF00020010>.
- (80) McElroy, D.; Zhang, W.; Cao, J.; Wu, R. Isolation of an Efficient Actin Promoter for Use in Rice Transformation. *Plant Cell* **1990**, *2* (2), 163–171. <https://doi.org/10.1105/tpc.2.2.163>.
- (81) Sainsbury, F.; Lomonossoff, G. P. Extremely High-Level and Rapid Transient Protein Production in Plants without the Use of Viral Replication. *Plant Physiol.* **2008**, *148* (3), 1212–1218. <https://doi.org/10.1104/pp.108.126284>.
- (82) Ali, S.; Kim, W.-C. A Fruitful Decade Using Synthetic Promoters in the Improvement of Transgenic Plants. *Front. Plant Sci.* **2019**, *10*. <https://doi.org/10.3389/fpls.2019.01433>.
- (83) Huang, D.; Kosentka, P. Z.; Liu, W. Synthetic Biology Approaches in Regulation of Targeted Gene Expression. *Curr. Opin. Plant Biol.* **2021**, *63*, 102036. <https://doi.org/10.1016/j.pbi.2021.102036>.
- (84) Belcher, M. S.; Vuu, K. M.; Zhou, A.; Mansoori, N.; Agosto Ramos, A.; Thompson, M. G.; Scheller, H. V.; Loqué, D.; Shih, P. M. Design of Orthogonal Regulatory Systems for Modulating Gene Expression in Plants. *Nat. Chem. Biol.* **2020**, 1–9. <https://doi.org/10.1038/s41589-020-0547-4>.
- (85) Cai, Y.-M.; Kallam, K.; Tidd, H.; Gendarini, G.; Salzman, A.; Patron, N. J. Rational Design of Minimal Synthetic Promoters for Plants. *Nucleic Acids Res.* **2020**. <https://doi.org/10.1093/nar/gkaa682>.
- (86) Gupta, D.; Dey, N.; Leelavathi, S.; Ranjan, R. Development of Efficient Synthetic Promoters Derived from Pararetrovirus Suitable for Translational Research. *Planta* **2021**, *253* (2), 42. <https://doi.org/10.1007/s00425-021-03565-9>.
- (87) Jores, T.; Tonnies, J.; Wrightsman, T.; Buckler, E. S.; Cuperus, J. T.; Fields, S.; Queitsch, C. Synthetic Promoter Designs Enabled by a Comprehensive Analysis of Plant Core Promoters. *Nat. Plants* **2021**, *7* (6), 842–855. <https://doi.org/10.1038/s41477-021-00932-y>.

- (88) Yang, Y.; Lee, J. H.; Poindexter, M. R.; Shao, Y.; Liu, W.; Lenaghan, S. C.; Ahkami, A. H.; Blumwald, E.; Stewart Jr, C. N. Rational Design and Testing of Abiotic Stress-Inducible Synthetic Promoters from Poplar Cis-Regulatory Elements. *Plant Biotechnol. J.* **2021**, *19* (7), 1354–1369. <https://doi.org/10.1111/pbi.13550>.
- (89) Damos, A. G.; Rosenthal, S. H.; Mason, H. S. 5' and 3' Untranslated Regions Strongly Enhance Performance of Geminiviral Replicons in *Nicotiana Benthamiana* Leaves. *Front. Plant Sci.* **2016**, *7*.
- (90) Peyret, H.; Brown, J. K. M.; Lomonossoff, G. P. Improving Plant Transient Expression through the Rational Design of Synthetic 5' and 3' Untranslated Regions. *Plant Methods* **2019**, *15* (1), 108. <https://doi.org/10.1186/s13007-019-0494-9>.
- (91) Damos, A. G.; Mason, H. S. Chimeric 3' Flanking Regions Strongly Enhance Gene Expression in Plants. *Plant Biotechnol. J.* **2018**, *16* (12), 1971–1982. <https://doi.org/10.1111/pbi.12931>.
- (92) Vanhercke, T.; Divi, U. K.; El Tahchy, A.; Liu, Q.; Mitchell, M.; Taylor, M. C.; Eastmond, P. J.; Bryant, F.; Mechanicos, A.; Blundell, C.; Zhi, Y.; Belide, S.; Shrestha, P.; Zhou, X.-R.; Ral, J.-P.; White, R. G.; Green, A.; Singh, S. P.; Petrie, J. R. Step Changes in Leaf Oil Accumulation via Iterative Metabolic Engineering. *Metab. Eng.* **2017**, *39*, 237–246. <https://doi.org/10.1016/j.ymben.2016.12.007>.
- (93) Wang, C.; Ding, D.; Yan, R.; Yu, X.; Li, W.; Li, M. A Novel Bi-Directional Promoter Cloned from Melon and Its Activity in Cucumber and Tobacco. *J. Plant Biol.* **2008**, *51* (2), 108–115. <https://doi.org/10.1007/BF03030719>.
- (94) Kourmpetli, S.; Lee, K.; Hemsley, R.; Rossignol, P.; Papageorgiou, T.; Drea, S. Bidirectional Promoters in Seed Development and Related Hormone/Stress Responses. *BMC Plant Biol.* **2013**, *13* (1), 187. <https://doi.org/10.1186/1471-2229-13-187>.
- (95) Zhang, C.; Gai, Y.; Zhu, Y.; Chen, X.; Jiang, X. Construction of a Bidirectional Promoter and Its Transient Expression in *Populus Tomentosa*. *Front. For. China* **2008**, *3* (1), 112–116. <https://doi.org/10.1007/s11461-008-0018-7>.
- (96) Zhang, C.; Gai, Y.; Wang, W.; Zhu, Y.; Chen, X.; Jiang, X. Construction and Analysis of a Plant Transformation Binary Vector PBDGG Harboring a Bi-Directional Promoter Fusing Dual Visible Reporter Genes. *J. Genet. Genomics* **2008**, *35* (4), 245–249. [https://doi.org/10.1016/S1673-8527\(08\)60034-X](https://doi.org/10.1016/S1673-8527(08)60034-X).
- (97) Verhounig, A.; Karcher, D.; Bock, R. Inducible Gene Expression from the Plastid Genome by a Synthetic Riboswitch. *Proc. Natl. Acad. Sci.* **2010**, *107* (14), 6204–6209. <https://doi.org/10.1073/pnas.0914423107>.

- (98) Agrawal, S.; Karcher, D.; Ruf, S.; Erban, A.; Hertle, A. P.; Kopka, J.; Bock, R. Riboswitch-Mediated Inducible Expression of an Astaxanthin Biosynthetic Operon in Plastids. *Plant Physiol.* **2022**, *188* (1), 637–652. <https://doi.org/10.1093/plphys/kiab428>.
- (99) Srivastava, A. K.; Lu, Y.; Zinta, G.; Lang, Z.; Zhu, J.-K. UTR Dependent Control of Gene Expression in Plants. *Trends Plant Sci.* **2018**, *23* (3), 248–259. <https://doi.org/10.1016/j.tplants.2017.11.003>.
- (100) Bernardes, W. S.; Menossi, M. Plant 3' Regulatory Regions From mRNA-Encoding Genes and Their Uses to Modulate Expression. *Front. Plant Sci.* **2020**, *11*.
- (101) Poliner, E.; Pulman, J. A.; Zienkiewicz, K.; Childs, K.; Benning, C.; Farré, E. M. A Toolkit for *Nannochloropsis Oceanica* CCMP1779 Enables Gene Stacking and Genetic Engineering of the Eicosapentaenoic Acid Pathway for Enhanced Long-chain Polyunsaturated Fatty Acid Production. *Plant Biotechnol. J.* **2018**, *16* (1), 298–309. <https://doi.org/10.1111/pbi.12772>.
- (102) Vogl, T.; Kickenweiz, T.; Pitzer, J.; Sturmberger, L.; Weninger, A.; Biggs, B. W.; Köhler, E.-M.; Baumschlager, A.; Fischer, J. E.; Hyden, P.; Wagner, M.; Baumann, M.; Borth, N.; Geier, M.; Ajikumar, P. K.; Glieder, A. Engineered Bidirectional Promoters Enable Rapid Multi-Gene Co-Expression Optimization. *Nat. Commun.* **2018**, *9* (1), 3589. <https://doi.org/10.1038/s41467-018-05915-w>.
- (103) de Felipe, P.; Hughes, L. E.; Ryan, M. D.; Brown, J. D. Co-Translational, Intraribosomal Cleavage of Polypeptides by the Foot-and-Mouth Disease Virus 2A Peptide. *J. Biol. Chem.* **2003**, *278* (13), 11441–11448. <https://doi.org/10.1074/jbc.M211644200>.
- (104) Sun, H.; Zhou, N.; Wang, H.; Huang, D.; Lang, Z. Processing and Targeting of Proteins Derived from Polyprotein with 2A and LP4/2A as Peptide Linkers in a Maize Expression System. *PLOS ONE* **2017**, *12* (3), e0174804. <https://doi.org/10.1371/journal.pone.0174804>.
- (105) Thimmappa, R.; Geisler, K.; Louveau, T.; O'Maille, P.; Osbourn, A. Triterpene Biosynthesis in Plants. *Annu. Rev. Plant Biol.* **2014**, *65*, 225–257. <https://doi.org/10.1146/annurev-arplant-050312-120229>.
- (106) Tsujimoto, Mitsumaru. A HIGHLY UNSATURATED HYDROCARBON IN SHARK LIVER OIL. *J. Ind. Eng. Chem.* **1916**, *8* (10), 889–896. <https://doi.org/10.1021/i500010a005>.
- (107) Spanova, M.; Daum, G. Squalene – Biochemistry, Molecular Biology, Process Biotechnology, and Applications. *Eur. J. Lipid Sci. Technol.* **2011**, *113* (11), 1299–1320. <https://doi.org/10.1002/ejlt.201100203>.

- (108) Tracy, N. I.; Crunkleton, D. W.; Price, G. L. Catalytic Cracking of Squalene to Gasoline-Range Molecules. *Biomass Bioenergy* **2011**, 35 (3), 1060–1065. <https://doi.org/10.1016/j.biombioe.2010.11.027>.
- (109) Zerbe, P.; Bohlmann, J. Enzymes for Synthetic Biology of Ambroxide-Related Diterpenoid Fragrance Compounds. In *Biotechnology of Isoprenoids*; Schrader, J., Bohlmann, J., Eds.; Advances in Biochemical Engineering/Biotechnology; Springer International Publishing: Cham, 2015; pp 427–447. https://doi.org/10.1007/10_2015_308.
- (110) Beltrán, G.; Bucheli, M. E.; Aguilera, M. P.; Belaj, A.; Jimenez, A. Squalene in Virgin Olive Oil: Screening of Variability in Olive Cultivars. *Eur. J. Lipid Sci. Technol.* **2016**, 118 (8), 1250–1253. <https://doi.org/10.1002/ejlt.201500295>.
- (111) Aharoni, A.; Giri, A. P.; Deuerlein, S.; Griepink, F.; Kogel, W.-J. de; Verstappen, F. W. A.; Verhoeven, H. A.; Jongsma, M. A.; Schwab, W.; Bouwmeester, H. J. Terpenoid Metabolism in Wild-Type and Transgenic *Arabidopsis* Plants. *Plant Cell* **2003**, 15 (12), 2866–2884. <https://doi.org/10.1105/tpc.016253>.
- (112) Besumbes, Ó.; Sauret-Güeto, S.; Phillips, M. A.; Imperial, S.; Rodríguez-Concepción, M.; Boronat, A. Metabolic Engineering of Isoprenoid Biosynthesis in *Arabidopsis* for the Production of Taxadiene, the First Committed Precursor of Taxol. *Biotechnol. Bioeng.* **2004**, 88 (2), 168–175. <https://doi.org/10.1002/bit.20237>.
- (113) Tholl, D. Biosynthesis and Biological Functions of Terpenoids in Plants. In *Biotechnology of Isoprenoids*; Schrader, J., Bohlmann, J., Eds.; Advances in Biochemical Engineering/Biotechnology; Springer International Publishing: Cham, 2015; pp 63–106. https://doi.org/10.1007/10_2014_295.
- (114) Cordoba, E.; Salmi, M.; León, P. Unravelling the Regulatory Mechanisms That Modulate the MEP Pathway in Higher Plants. *J. Exp. Bot.* **2009**, 60 (10), 2933–2943. <https://doi.org/10.1093/jxb/erp190>.
- (115) Banerjee, A.; D. Sharkey, T. Methylerythritol 4-Phosphate (MEP) Pathway Metabolic Regulation. *Nat. Prod. Rep.* **2014**, 31 (8), 1043–1055. <https://doi.org/10.1039/C3NP70124G>.
- (116) Ikram, N. K. B. K.; Zhan, X.; Pan, X.-W.; King, B. C.; Simonsen, H. T. Stable Heterologous Expression of Biologically Active Terpenoids in Green Plant Cells. *Front. Plant Sci.* **2015**, 6. <https://doi.org/10.3389/fpls.2015.00129>.
- (117) Rodríguez-Concepción, M.; Boronat, A. Breaking New Ground in the Regulation of the Early Steps of Plant Isoprenoid Biosynthesis. *Curr. Opin. Plant Biol.* **2015**, 25, 17–22. <https://doi.org/10.1016/j.pbi.2015.04.001>.
- (118) Banerjee, A.; Wu, Y.; Banerjee, R.; Li, Y.; Yan, H.; Sharkey, T. D. Feedback Inhibition of Deoxy-d-Xylulose-5-Phosphate Synthase Regulates the

- Methylerythritol 4-Phosphate Pathway. *J. Biol. Chem.* **2013**, 288 (23), 16926–16936. <https://doi.org/10.1074/jbc.M113.464636>.
- (119) Kwan, T. A.; Kwan, S. E.; Peccia, J.; Zimmerman, J. B. Selectively Biorefining Astaxanthin and Triacylglycerol Co-Products from Microalgae with Supercritical Carbon Dioxide Extraction. *Bioresour. Technol.* **2018**, 269, 81–88. <https://doi.org/10.1016/j.biortech.2018.08.081>.
- (120) Ren, Y.; Deng, J.; Huang, J.; Wu, Z.; Yi, L.; Bi, Y.; Chen, F. Using Green Alga *Haematococcus Pluvialis* for Astaxanthin and Lipid Co-Production: Advances and Outlook. *Bioresour. Technol.* **2021**, 340, 125736. <https://doi.org/10.1016/j.biortech.2021.125736>.
- (121) Vieler, A.; Brubaker, S. B.; Vick, B.; Benning, C. A Lipid Droplet Protein of *Nannochloropsis* with Functions Partially Analogous to Plant Oleosins. *Plant Physiol.* **2012**, 158 (4), 1562–1569. <https://doi.org/10.1104/pp.111.193029>.
- (122) Lee, D. W.; Lee, S.; Lee, G.; Lee, K. H.; Kim, S.; Cheong, G.-W.; Hwang, I. Functional Characterization of Sequence Motifs in the Transit Peptide of *Arabidopsis* Small Subunit of Rubisco. *Plant Physiol.* **2006**, 140 (2), 466–483. <https://doi.org/10.1104/pp.105.074575>.
- (123) Arrivault, S.; Guenther, M.; Ivakov, A.; Feil, R.; Vosloh, D.; Dongen, J. T. V.; Sulpice, R.; Stitt, M. Use of Reverse-Phase Liquid Chromatography, Linked to Tandem Mass Spectrometry, to Profile the Calvin Cycle and Other Metabolic Intermediates in *Arabidopsis* Rosettes at Different Carbon Dioxide Concentrations. *Plant J.* **2009**, 59 (5), 826–839. <https://doi.org/10.1111/j.1365-313X.2009.03902.x>.
- (124) Hu, T.; Zhou, J.; Tong, Y.; Su, P.; Li, X.; Liu, Y.; Liu, N.; Wu, X.; Zhang, Y.; Wang, J.; Gao, L.; Tu, L.; Lu, Y.; Jiang, Z.; Zhou, Y. J.; Gao, W.; Huang, L. Engineering Chimeric Diterpene Synthases and Isoprenoid Biosynthetic Pathways Enables High-Level Production of Miltiradiene in Yeast. *Metab. Eng.* **2020**, 60, 87–96. <https://doi.org/10.1016/j.ymben.2020.03.011>.
- (125) Huang, D.; Yao, Y.; Zhang, H.; Mei, Z.; Wang, R.; Feng, L.; Liu, B. Directed Optimization of a Newly Identified Squalene Synthase from *Mortierella Alpine* Based on Sequence Truncation and Site-Directed Mutagenesis. *J. Ind. Microbiol. Biotechnol.* **2015**, 42 (10), 1341–1352. <https://doi.org/10.1007/s10295-015-1668-8>.
- (126) Athanasakoglou, A.; Grypioti, E.; Michailidou, S.; Ignea, C.; Makris, A. M.; Kalantidis, K.; Massé, G.; Argiriou, A.; Verret, F.; Kampranis, S. C. Isoprenoid Biosynthesis in the Diatom *Haslea Ostrearia*. *New Phytol.* **2019**, 222 (1), 230–243. <https://doi.org/10.1111/nph.15586>.
- (127) Ignea, C.; Cvetkovic, I.; Loupassaki, S.; Kefalas, P.; Johnson, C. B.; Kampranis, S. C.; Makris, A. M. Improving Yeast Strains Using Recyclable Integration

- Cassettes, for the Production of Plant Terpenoids. *Microb. Cell Factories* **2011**, 10 (1), 1–18. <https://doi.org/10.1186/1475-2859-10-4>.
- (128) Zienkiewicz, K.; Zienkiewicz, A. Degradation of Lipid Droplets in Plants and Algae—Right Time, Many Paths, One Goal. *Front. Plant Sci.* **2020**, 11.
- (129) Peramuna, A.; Bae, H.; López, C. Q.; Fromberg, A.; Petersen, B.; Simonsen, H. T. Connecting Moss Lipid Droplets to Patchoulol Biosynthesis. *PLOS ONE* **2020**, 15 (12), e0243620. <https://doi.org/10.1371/journal.pone.0243620>.
- (130) Grimberg, Å.; Carlsson, A. S.; Marttila, S.; Bhalerao, R.; Hofvander, P. Transcriptional Transitions in *Nicotiana Benthamiana* Leaves upon Induction of Oil Synthesis by WRINKLED1 Homologs from Diverse Species and Tissues. *BMC Plant Biol.* **2015**, 15 (1), 192. <https://doi.org/10.1186/s12870-015-0579-1>.
- (131) Ma, W.; Kong, Q.; Grix, M.; Mantyla, J. J.; Yang, Y.; Benning, C.; Ohlrogge, J. B. Deletion of a C-Terminal Intrinsically Disordered Region of WRINKLED1 Affects Its Stability and Enhances Oil Accumulation in *Arabidopsis*. *Plant J.* **2015**, 83 (5), 864–874. <https://doi.org/10.1111/tpj.12933>.
- (132) Kavanagh, K. L.; Guo, K.; Dunford, J. E.; Wu, X.; Knapp, S.; Ebetino, F. H.; Rogers, M. J.; Russell, R. G. G.; Oppermann, U. The Molecular Mechanism of Nitrogen-Containing Bisphosphonates as Antiosteoporosis Drugs. *Proc. Natl. Acad. Sci.* **2006**, 103 (20), 7829–7834. <https://doi.org/10.1073/pnas.0601643103>.
- (133) François, I. E. J. A.; Van Hemelrijck, W.; Aerts, A. M.; Wouters, P. F. J.; Proost, P.; Broekaert, W. F.; Cammue, B. P. A. Processing in *Arabidopsis Thaliana* of a Heterologous Polyprotein Resulting in Differential Targeting of the Individual Plant Defensins. *Plant Sci.* **2004**, 166 (1), 113–121. <https://doi.org/10.1016/j.plantsci.2003.09.001>.
- (134) Vidi, P.-A.; Kanwischer, M.; Baginsky, S.; Austin, J. R.; Csucs, G.; Dörmann, P.; Kessler, F.; Bréhélin, C. Tocopherol Cyclase (VTE1) Localization and Vitamin E Accumulation in Chloroplast Plastoglobule Lipoprotein Particles. *J. Biol. Chem.* **2006**, 281 (16), 11225–11234. <https://doi.org/10.1074/jbc.M511939200>.
- (135) Espinoza-Corral, R.; Schwenkert, S.; Lundquist, P. K. Molecular Changes of *Arabidopsis Thaliana* Plastoglobules Facilitate Thylakoid Membrane Remodeling under High Light Stress. *Plant J. Cell Mol. Biol.* **2021**. <https://doi.org/10.1111/tpj.15253>.
- (136) Dellas, N.; Thomas, S. T.; Manning, G.; Noel, J. P. Discovery of a Metabolic Alternative to the Classical Mevalonate Pathway. *eLife* **2013**, 2, e00672. <https://doi.org/10.7554/eLife.00672>.

- (137) Valachovic, M.; Garaiova, M.; Holic, R.; Hapala, I. Squalene Is Lipotoxic to Yeast Cells Defective in Lipid Droplet Biogenesis. *Biochem. Biophys. Res. Commun.* **2016**, *469* (4), 1123–1128. <https://doi.org/10.1016/j.bbrc.2015.12.050>.
- (138) Csáky, Z.; Garaiová, M.; Kodedová, M.; Valachovič, M.; Sychrová, H.; Hapala, I. Squalene Lipotoxicity in a Lipid Droplet-Less Yeast Mutant Is Linked to Plasma Membrane Dysfunction. *Yeast* **2020**, *37* (1), 45–62. <https://doi.org/10.1002/yea.3454>.
- (139) Melis, A. Carbon Partitioning in Photosynthesis. *Curr. Opin. Chem. Biol.* **2013**, *17* (3), 453–456. <https://doi.org/10.1016/j.cbpa.2013.03.010>.
- (140) Sainsbury, F.; Thuenemann, E. C.; Lomonossoff, G. P. PEAQ: Versatile Expression Vectors for Easy and Quick Transient Expression of Heterologous Proteins in Plants. *Plant Biotechnol. J.* **2009**, *7* (7), 682–693. <https://doi.org/10.1111/j.1467-7652.2009.00434.x>.
- (141) Hoagland, D. R.; Arnon, D. I. Growing plants without soil by the water-culture method. *Grow. Plants Soil Water-Cult. Method* **1938**.
- (142) Farquhar, G. D.; von Caemmerer, S. Modelling of Photosynthetic Response to Environmental Conditions. In *Physiological Plant Ecology II: Water Relations and Carbon Assimilation*; Lange, O. L., Nobel, P. S., Osmond, C. B., Ziegler, H., Eds.; Encyclopedia of Plant Physiology; Springer: Berlin, Heidelberg, 1982; pp 549–587. https://doi.org/10.1007/978-3-642-68150-9_17.
- (143) Sharkey, T. D.; Bernacchi, C. J.; Farquhar, G. D.; Singsaas, E. L. Fitting Photosynthetic Carbon Dioxide Response Curves for C₃ Leaves. *Plant Cell Environ.* **2007**, *30* (9), 1035–1040. <https://doi.org/10.1111/j.1365-3040.2007.01710.x>.
- (144) Sharkey, T. D. What Gas Exchange Data Can Tell Us about Photosynthesis. *Plant Cell Environ.* **2016**, *39* (6), 1161–1163. <https://doi.org/10.1111/pce.12641>.
- (145) Gregory, L. M.; McClain, A. M.; Kramer, D. M.; Pardo, J. D.; Smith, K. E.; Tessmer, O. L.; Walker, B. J.; Ziccardi, L. G.; Sharkey, T. D. The Triose Phosphate Utilization Limitation of Photosynthetic Rate: Out of Global Models but Important for Leaf Models. *Plant Cell Environ.* **2021**, *44* (10), 3223–3226. <https://doi.org/10.1111/pce.14153>.
- (146) Bohlmann, J.; Keeling, C. I. Terpenoid Biomaterials. *Plant J.* **2008**, *54* (4), 656–669. <https://doi.org/10.1111/j.1365-313X.2008.03449.x>.
- (147) Zhou, S.; Runge, T.; Karlen, S. D.; Ralph, J.; Gonzales-Vigil, E.; Mansfield, S. D. Chemical Pulping Advantages of Zip-Lignin Hybrid Poplar. *ChemSusChem* **2017**, *10* (18), 3565–3573. <https://doi.org/10.1002/cssc.201701317>.

- (148) Bhalla, A.; Bansal, N.; Pattathil, S.; Li, M.; Shen, W.; Particka, C. A.; Karlen, S. D.; Phongpreecha, T.; Semaan, R. R.; Gonzales-Vigil, E.; Ralph, J.; Mansfield, S. D.; Ding, S.-Y.; Hodge, D. B.; Hegg, E. L. Engineered Lignin in Poplar Biomass Facilitates Cu-Catalyzed Alkaline-Oxidative Pretreatment. *ACS Sustain. Chem. Eng.* **2018**, 6 (3), 2932–2941. <https://doi.org/10.1021/acssuschemeng.7b02067>.
- (149) An, Y.; Liu, Y.; Liu, Y.; Lu, M.; Kang, X.; Mansfield, S. D.; Zeng, W.; Zhang, J. Opportunities and Barriers for Biofuel and Bioenergy Production from Poplar. *GCB Bioenergy* **2021**, 13 (6), 905–913. <https://doi.org/10.1111/gcbb.12829>.
- (150) Labrecque, M.; Teodorescu, T. I. Field Performance and Biomass Production of 12 Willow and Poplar Clones in Short-Rotation Coppice in Southern Quebec (Canada). *Biomass Bioenergy* **2005**, 29 (1), 1–9. <https://doi.org/10.1016/j.biombioe.2004.12.004>.
- (151) Yevtushenko, D. P.; Misra, S. Efficient *Agrobacterium*-Mediated Transformation of Commercial Hybrid Poplar *Populus Nigra* L. × *P. Maximowiczii* A. Henry. *Plant Cell Rep.* **2010**, 29 (3), 211–221. <https://doi.org/10.1007/s00299-009-0806-z>.
- (152) Ko, J.-H.; Kim, H.-T.; Hwang, I.; Han, K.-H. Tissue-Type-Specific Transcriptome Analysis Identifies Developing Xylem-Specific Promoters in Poplar. *Plant Biotechnol. J.* **2012**, 10 (5), 587–596. <https://doi.org/10.1111/j.1467-7652.2012.00690.x>.
- (153) Han, X.; Ma, S.; Kong, X.; Takano, T.; Liu, S. Efficient *Agrobacterium*-Mediated Transformation of Hybrid Poplar *Populus Davidiana* Dode × *Populus Bollena* Lauche. *Int. J. Mol. Sci.* **2013**, 14 (2), 2515–2528. <https://doi.org/10.3390/ijms14022515>.
- (154) Vining, K.; Pomraning, K. R.; Wilhelm, L. J.; Ma, C.; Pellegrini, M.; Di, Y.; Mockler, T. C.; Freitag, M.; Strauss, S. H. Methylome Reorganization during in Vitro Dedifferentiation and Regeneration of *Populus Trichocarpa*. *BMC Plant Biol.* **2013**, 13 (1), 92. <https://doi.org/10.1186/1471-2229-13-92>.
- (155) Zhang, W.; Wang, Y.; Diao, S.; Zhong, S.; Wu, S.; Wang, L.; Su, X.; Zhang, B. Assessment of Epigenetic and Phenotypic Variation in *Populus Nigra* Regenerated via Sequential Regeneration. *Front. Plant Sci.* **2021**, 12.
- (156) Kong, Q.; Ma, W.; Yang, H.; Ma, G.; Mantyla, J. J.; Benning, C. The *Arabidopsis* WRINKLED1 Transcription Factor Affects Auxin Homeostasis in Roots. *J. Exp. Bot.* **2017**, 68 (16), 4627–4634. <https://doi.org/10.1093/jxb/erx275>.
- (157) Kong, Q.; Low, P. M.; Lim, A. R. Q.; Yang, Y.; Yuan, L.; Ma, W. Functional Antagonism of WRI1 and TCP20 Modulates GH3.3 Expression to Maintain Auxin Homeostasis in Roots. *Plants* **2022**, 11 (3), 454. <https://doi.org/10.3390/plants11030454>.

- (158) Bannoud, F.; Bellini, C. Adventitious Rooting in *Populus* Species: Update and Perspectives. *Front. Plant Sci.* **2021**, *12*.
- (159) Pollastri, S.; Baccelli, I.; Loreto, F. Isoprene: An Antioxidant Itself or a Molecule with Multiple Regulatory Functions in Plants? *Antioxidants* **2021**, *10* (5), 684. <https://doi.org/10.3390/antiox10050684>.
- (160) Behnke, K.; Ehlting, B.; Teuber, M.; Bauerfeind, M.; Louis, S.; Hänsch, R.; Polle, A.; Bohlmann, J.; Schnitzler, J.-P. Transgenic, Non-Isoprene Emitting Poplars Don't like It Hot. *Plant J.* **2007**, *51* (3), 485–499. <https://doi.org/10.1111/j.1365-313X.2007.03157.x>.
- (161) Behnke, K.; Grote, R.; Brüggemann, N.; Zimmer, I.; Zhou, G.; Elobeid, M.; Janz, D.; Polle, A.; Schnitzler, J.-P. Isoprene Emission-Free Poplars – a Chance to Reduce the Impact from Poplar Plantations on the Atmosphere. *New Phytol.* **2012**, *194* (1), 70–82. <https://doi.org/10.1111/j.1469-8137.2011.03979.x>.
- (162) Monson, R. K.; Winkler, B.; Rosenstiel, T. N.; Block, K.; Merl-Pham, J.; Strauss, S. H.; Ault, K.; Maxfield, J.; Moore, D. J. P.; Trahan, N. A.; Neice, A. A.; Shiach, I.; Barron-Gafford, G. A.; Ibsen, P.; McCorkel, J. T.; Bernhardt, J.; Schnitzler, J.-P. High Productivity in Hybrid-Poplar Plantations without Isoprene Emission to the Atmosphere. *Proc. Natl. Acad. Sci.* **2020**, *117* (3), 1596–1605. <https://doi.org/10.1073/pnas.1912327117>.
- (163) Macdonald, C.; Soll, J. Shark Conservation Risks Associated with the Use of Shark Liver Oil in SARS-CoV-2 Vaccine Development. *bioRxiv* October 15, 2020, p 2020.10.14.338053. <https://doi.org/10.1101/2020.10.14.338053>.
- (164) Guenther, A. B.; Hills, A. J. Eddy Covariance Measurement of Isoprene Fluxes. *J. Geophys. Res. Atmospheres* **1998**, *103* (D11), 13145–13152. <https://doi.org/10.1029/97JD03283>.
- (165) Woods, D. R. *Rules of Thumb in Engineering Practice*; John Wiley & Sons, 2007.
- (166) Pouvreau, B.; Vanhercke, T.; Singh, S. From Plant Metabolic Engineering to Plant Synthetic Biology: The Evolution of the Design/Build/Test/Learn Cycle. *Plant Sci.* **2018**, *273*, 3–12. <https://doi.org/10.1016/j.plantsci.2018.03.035>.
- (167) Ikram, N. K. B. K.; Simonsen, H. T. A Review of Biotechnological Artemisinin Production in Plants. *Front. Plant Sci.* **2017**, *8*.
- (168) Mohan, C.; Jayanarayanan, A. N.; Narayanan, S. Construction of a Novel Synthetic Root-Specific Promoter and Its Characterization in Transgenic Tobacco Plants. *3 Biotech* **2017**, *7* (4), 234. <https://doi.org/10.1007/s13205-017-0872-9>.
- (169) Ochoa-Fernandez, R.; Abel, N. B.; Wieland, F.-G.; Schlegel, J.; Koch, L.-A.; Miller, J. B.; Engesser, R.; Giuriani, G.; Brandl, S. M.; Timmer, J.; Weber, W.;

- Ott, T.; Simon, R.; Zurbriggen, M. D. Optogenetic Control of Gene Expression in Plants in the Presence of Ambient White Light. *Nat. Methods* **2020**, *17* (7), 717–725. <https://doi.org/10.1038/s41592-020-0868-y>.
- (170) Pan, C.; Wu, X.; Markel, K.; Malzahn, A. A.; Kundagrami, N.; Sretenovic, S.; Zhang, Y.; Cheng, Y.; Shih, P. M.; Qi, Y. CRISPR–Act3.0 for Highly Efficient Multiplexed Gene Activation in Plants. *Nat. Plants* **2021**, *7* (7), 942–953. <https://doi.org/10.1038/s41477-021-00953-7>.
- (171) Dhadi, S. R.; Krom, N.; Ramakrishna, W. Genome-Wide Comparative Analysis of Putative Bidirectional Promoters from Rice, *Arabidopsis* and *Populus*. *Gene* **2009**, *429* (1), 65–73. <https://doi.org/10.1016/j.gene.2008.09.034>.
- (172) Mitra, A.; Han, J.; Zhang, Z. J.; Mitra, A. The Intergenic Region of *Arabidopsis Thaliana* Cab1 and Cab2 Divergent Genes Functions as a Bidirectional Promoter. *Planta* **2009**, *229* (5), 1015–1022. <https://doi.org/10.1007/s00425-008-0859-1>.
- (173) Yang, J.; Wang, X.; Hasi, A.; Wang, Z. Structural and Functional Analysis of a Bidirectional Promoter from *Gossypium Hirsutum* in *Arabidopsis*. *Int. J. Mol. Sci.* **2018**, *19* (11), 3291. <https://doi.org/10.3390/ijms19113291>.
- (174) Alok, A.; Tiwari, S.; Tuli, R. A Bidirectional Promoter from Papaya Leaf Crumple Virus Functions in Both Monocot and Dicot Plants. *Physiol. Mol. Plant Pathol.* **2019**, *108*, 101423. <https://doi.org/10.1016/j.pmpp.2019.101423>.
- (175) Lis, M.; Walther, D. The Orientation of Transcription Factor Binding Site Motifs in Gene Promoter Regions: Does It Matter? *BMC Genomics* **2016**, *17*. <https://doi.org/10.1186/s12864-016-2549-x>.
- (176) Tuskan, G. A.; DiFazio, S.; Jansson, S.; Bohlmann, J.; Grigoriev, I.; Hellsten, U.; Putnam, N.; Ralph, S.; Rombauts, S.; Salamov, A.; Schein, J.; Sterck, L.; Aerts, A.; Bhalarao, R. R.; Bhalarao, R. P.; Blaudez, D.; Boerjan, W.; Brun, A.; Brunner, A.; Busov, V.; Campbell, M.; Carlson, J.; Chalot, M.; Chapman, J.; Chen, G.-L.; Cooper, D.; Coutinho, P. M.; Couturier, J.; Covert, S.; Cronk, Q.; Cunningham, R.; Davis, J.; Degroove, S.; Déjardin, A.; dePamphilis, C.; Detter, J.; Dirks, B.; Dubchak, I.; Duplessis, S.; Ehlting, J.; Ellis, B.; Gendler, K.; Goodstein, D.; Gribskov, M.; Grimwood, J.; Groover, A.; Gunter, L.; Hamberger, B.; Heinze, B.; Helariutta, Y.; Henrissat, B.; Holligan, D.; Holt, R.; Huang, W.; Islam-Faridi, N.; Jones, S.; Jones-Rhoades, M.; Jorgensen, R.; Joshi, C.; Kangasjärvi, J.; Karlsson, J.; Kelleher, C.; Kirkpatrick, R.; Kirst, M.; Kohler, A.; Kalluri, U.; Larimer, F.; Leebens-Mack, J.; Leplé, J.-C.; Locascio, P.; Lou, Y.; Lucas, S.; Martin, F.; Montanini, B.; Napoli, C.; Nelson, D. R.; Nelson, C.; Nieminen, K.; Nilsson, O.; Pereda, V.; Peter, G.; Philippe, R.; Pilate, G.; Poliakov, A.; Razumovskaya, J.; Richardson, P.; Rinaldi, C.; Ritland, K.; Rouzé, P.; Ryaboy, D.; Schmutz, J.; Schrader, J.; Segerman, B.; Shin, H.; Siddiqui, A.; Sterky, F.; Terry, A.; Tsai, C.-J.; Uberbacher, E.; Unneberg, P.; Vahala, J.; Wall, K.;

- Wessler, S.; Yang, G.; Yin, T.; Douglas, C.; Marra, M.; Sandberg, G.; Van de Peer, Y.; Rokhsar, D. The Genome of Black Cottonwood, *Populus Trichocarpa* (Torr. & Gray). *Science* **2006**, *313* (5793), 1596–1604. <https://doi.org/10.1126/science.1128691>.
- (177) Cheng, C.-Y.; Krishnakumar, V.; Chan, A. P.; Thibaud-Nissen, F.; Schobel, S.; Town, C. D. Araport11: A Complete Reannotation of the *Arabidopsis Thaliana* Reference Genome. *Plant J.* **2017**, *89* (4), 789–804. <https://doi.org/10.1111/tpj.13415>.
- (178) Rizzo, M. A.; Springer, G.; Segawa, K.; Zipfel, W. R.; Piston, D. W. Optimization of Pairings and Detection Conditions for Measurement of FRET between Cyan and Yellow Fluorescent Proteins. *Microsc. Microanal.* **2006**, *12* (3), 238–254. <https://doi.org/10.1017/S1431927606060235>.
- (179) Spiess, E.; Bestvater, F.; Heckel-Pompey, A.; Toth, K.; Hacker, M.; Stobrawa, G.; Feurer, T.; Wotzlaw, C.; Berchner-Pfannschmidt, U.; Porwol, T.; Acker, H. Two-Photon Excitation and Emission Spectra of the Green Fluorescent Protein Variants ECFP, EGFP and EYFP. *J. Microsc.* **2005**, *217* (3), 200–204. <https://doi.org/10.1111/j.1365-2818.2005.01437.x>.
- (180) Shaner, N. C.; Lin, M. Z.; McKeown, M. R.; Steinbach, P. A.; Hazelwood, K. L.; Davidson, M. W.; Tsien, R. Y. Improving the Photostability of Bright Monomeric Orange and Red Fluorescent Proteins. *Nat. Methods* **2008**, *5* (6), 545–551. <https://doi.org/10.1038/nmeth.1209>.
- (181) Gandhi, R.; Maheshwari, S. C.; Khurana, P. Transient Gene Expression and Influence of Promoters on Foreign Gene Expression in *Arabidopsis Thaliana*. *Vitro Cell. Dev. Biol. - Plant* **1999**, *35* (3), 232–237. <https://doi.org/10.1007/s11627-999-0084-z>.
- (182) Hong, J. K.; Suh, E. J.; Kwon, S.-J.; Lee, S. B.; Kim, J. A.; Lee, S. I.; Lee, Y.-H. Promoter of Chrysanthemum Actin Confers High-Level Constitutive Gene Expression in Arabidopsis and Chrysanthemum. *Sci. Hortic.* **2016**, *211*, 8–18. <https://doi.org/10.1016/j.scienta.2016.08.006>.
- (183) Schledzewski, K.; Mendel, R. R. Quantitative Transient Gene Expression: Comparison of the Promoters for Maize Polyubiquitin1, Rice Actin1, Maize-Derived Emu And CaMV 35S in Cells of Barley, Maize and Tobacco. *Transgenic Res.* **1994**, *3* (4), 249–255. <https://doi.org/10.1007/BF02336778>.
- (184) Mo, G. C. H.; Posner, C.; Rodriguez, E. A.; Sun, T.; Zhang, J. A Rationally Enhanced Red Fluorescent Protein Expands the Utility of FRET Biosensors. *Nat. Commun.* **2020**, *11* (1), 1848. <https://doi.org/10.1038/s41467-020-15687-x>.
- (185) F. de Felippes, F.; McHale, M.; Doran, R. L.; Roden, S.; Eamens, A. L.; Finnegan, E. J.; Waterhouse, P. M. The Key Role of Terminators on the

- Expression and Post-Transcriptional Gene Silencing of Transgenes. *Plant J.* **2020**, *104* (1), 96–112. <https://doi.org/10.1111/tpj.14907>.
- (186) Chow, C.-N.; Lee, T.-Y.; Hung, Y.-C.; Li, G.-Z.; Tseng, K.-C.; Liu, Y.-H.; Kuo, P.-L.; Zheng, H.-Q.; Chang, W.-C. PlantPAN3.0: A New and Updated Resource for Reconstructing Transcriptional Regulatory Networks from ChIP-Seq Experiments in Plants. *Nucleic Acids Res.* **2019**, *47* (Database issue), D1155–D1163. <https://doi.org/10.1093/nar/gky1081>.
 - (187) Liu, X.; Yang, W.; Mu, B.; Li, S.; Li, Y.; Zhou, X.; Zhang, C.; Fan, Y.; Chen, R. Engineering of ‘Purple Embryo Maize’ with a Multigene Expression System Derived from a Bidirectional Promoter and Self-cleaving 2A Peptides. *Plant Biotechnol. J.* **2018**, *16* (6), 1107–1109. <https://doi.org/10.1111/pbi.12883>.
 - (188) Goodstein, D. M.; Shu, S.; Howson, R.; Neupane, R.; Hayes, R. D.; Fazo, J.; Mitros, T.; Dirks, W.; Hellsten, U.; Putnam, N.; Rokhsar, D. S. Phytozome: A Comparative Platform for Green Plant Genomics. *Nucleic Acids Res.* **2012**, *40* (D1), D1178–D1186. <https://doi.org/10.1093/nar/gkr944>.
 - (189) Robić, G.; Lacorte, C.; Miranda, E. A. Fluorometric Quantification of Green Fluorescent Protein in Tobacco Leaf Extracts. *Anal. Biochem.* **2009**, *392* (1), 8–11. <https://doi.org/10.1016/j.ab.2009.05.016>.
 - (190) Zhang, Y.; Fernie, A. R. Metabolons, Enzyme–Enzyme Assemblies That Mediate Substrate Channeling, and Their Roles in Plant Metabolism. *Plant Commun.* **2021**, *2* (1), 100081. <https://doi.org/10.1016/j.xplc.2020.100081>.
 - (191) Vanhercke, T.; Dyer, J. M.; Mullen, R. T.; Kilaru, A.; Rahman, Md. M.; Petrie, J. R.; Green, A. G.; Yurchenko, O.; Singh, S. P. Metabolic Engineering for Enhanced Oil in Biomass. *Prog. Lipid Res.* **2019**, *74*, 103–129. <https://doi.org/10.1016/j.plipres.2019.02.002>.
 - (192) Du, Y.; Fu, X.; Chu, Y.; Wu, P.; Liu, Y.; Ma, L.; Tian, H.; Zhu, B. Biosynthesis and the Roles of Plant Sterols in Development and Stress Responses. *Int. J. Mol. Sci.* **2022**, *23* (4), 2332. <https://doi.org/10.3390/ijms23042332>.
 - (193) Wei, H.; Movahedi, A.; Xu, C.; Sun, W.; Li, L.; Wang, P.; Li, D.; Zhuge, Q. Overexpression of PtHMGR Enhances Drought and Salt Tolerance of Poplar. *Ann. Bot.* **2020**, *125* (5), 785–803. <https://doi.org/10.1093/aob/mcz158>.
 - (194) Xun, H.; Zhang, X.; Yu, J.; Pang, J.; Wang, S.; Liu, B.; Dong, Y.; Jiang, L.; Guo, D. Analysis of Expression Characteristics of Soybean Leaf and Root Tissue-Specific Promoters in *Arabidopsis* and Soybean. *Transgenic Res.* **2021**, *30* (6), 799–810. <https://doi.org/10.1007/s11248-021-00266-7>.
 - (195) Ng, R. T. L.; Maravelias, C. T. Economic and Energetic Analysis of Biofuel Supply Chains. *Appl. Energy* **2017**, *205* (May), 1571–1582. <https://doi.org/10.1016/j.apenergy.2017.08.161>.

- (196) Cheng, M. H.; Rosentrater, K. A. Economic Feasibility Analysis of Soybean Oil Production by Hexane Extraction. *Ind. Crops Prod.* **2017**, *108* (February), 775–785. <https://doi.org/10.1016/j.indcrop.2017.07.036>.
- (197) Humbird, D.; Davis, R.; Tao, L.; Kinchin, C.; Hsu, D.; Aden, A.; Schoen, P.; Lukas, J.; Olthof, B.; Worley, M.; Sexton, D.; Dudgeon, D. Process Design and Economics for Conversion of Lignocellulosic Biomass to Ethanol. *NREL Tech. Rep. NRELTP-5100-51400* **2011**, 303 (May 2011), 275–3000.