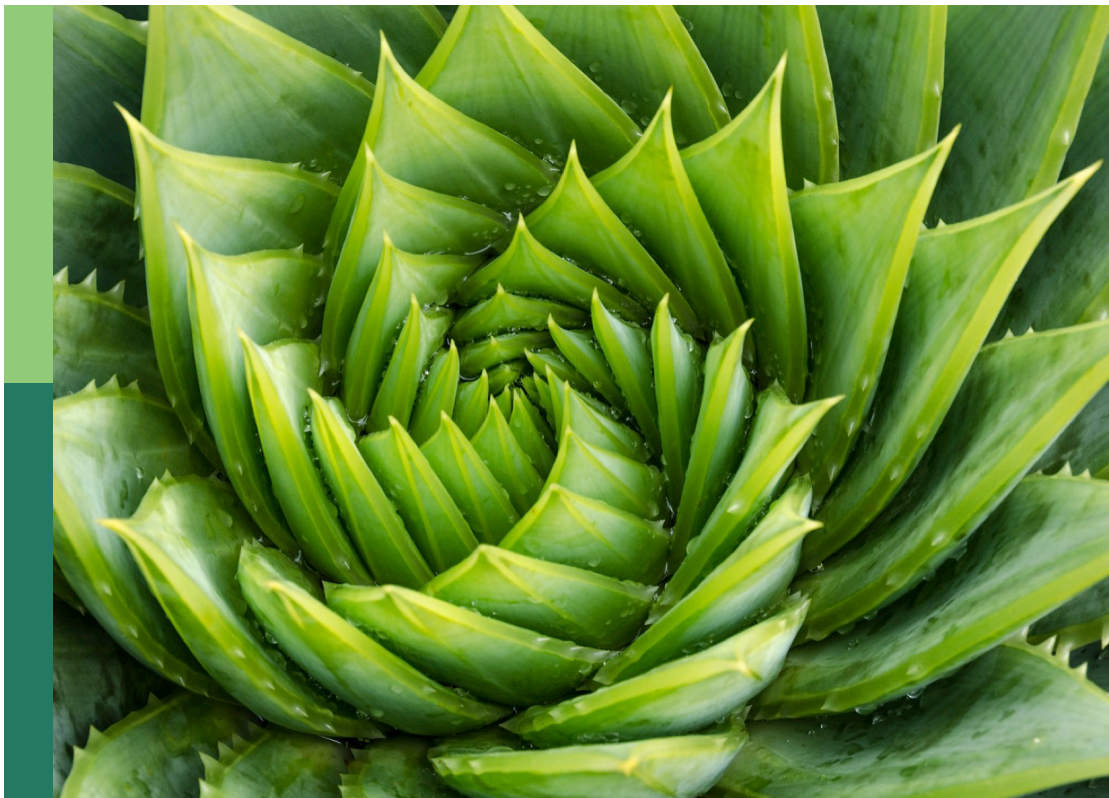
frontiers | Research Topics

Metabolic engineering of valuable compounds in photosynthetic organisms

Edited by
Zhi-Yan Du, Wajid Waheed Bhat, Guoyin Kai,
Inna Khozin-Goldberg, Xiao-Hong Yu,
Agnieszka Zienkiewicz and Krzysztof Zienkiewicz

Published in
Frontiers in Plant Science



ISSN 1664-8714
ISBN 978-2-8325-3513-4
DOI 10.3389/978-2-8325-3513-4

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Metabolic engineering of valuable compounds in photosynthetic organisms

Topic editors

Zhi-Yan Du — University of Hawaii at Manoa, United States
Wajid Waheed Bhat — Indian Institute of Integrative Medicine (CSIR), India
Guoyin Kai — Zhejiang Chinese Medical University, China
Inna Khozin-Goldberg — Ben-Gurion University of the Negev, Israel
Xiao-Hong Yu — Brookhaven National Laboratory (DOE), United States
Agnieszka Zienkiewicz — Nicolaus Copernicus University in Toruń, Poland
Krzysztof Zienkiewicz — Nicolaus Copernicus University in Toruń, Poland

Citation

Du, Z.-Y., Bhat, W. W., Kai, G., Khozin-Goldberg, I., Yu, X.-H., Zienkiewicz, A., Zienkiewicz, K., eds. (2023). *Metabolic engineering of valuable compounds in photosynthetic organisms*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-3513-4

Table of contents

frontiers | Frontiers in Plant Science

TYPE Editorial
PUBLISHED 05 September 2023
DOI 10.3389/fpls.2023.1260454

OPEN ACCESS

EDITED AND REVIEWED BY
Mark Blyth,
University of East Anglia, United Kingdom

*CORRESPONDENCE
Zhi-Yan Du
duz@hawaii.edu

RECEIVED 17 July 2023
ACCEPTED 22 August 2023
PUBLISHED 05 September 2023

CITATION
Du Z-Y, Bhat WW, Kai G, Khozin-Goldberg I, Yu X-H, Zienkiewicz A and Zienkiewicz K (2023) Editorial: Metabolic engineering of valuable compounds in photosynthetic organisms.
Front. Plant Sci. 14:1260454.
doi: 10.3389/fpls.2023.1260454

Editorial: Metabolic engineering of valuable compounds in photosynthetic organisms

Zhi-Yan Du[1]*, Wajid Waheed Bhat[2], Guoyin Kai[3], Inna Khozin-Goldberg[4], Xiao-Hong Yu[5], Agnieszka Zienkiewicz[6] and Krzysztof Zienkiewicz[6]

[1]Department of Molecular Biosciences and Bioengineering, University of Hawaii at Manoa, Honolulu, HI, United States, [2]Plant Biotechnology and Agrotechnology Division, Council of Scientific and Industrial Research-Indian Institute of Integrative Medicine (CSIR-IIIM), Jammu, India, [3]Laboratory of Medicinal Plant Biotechnology, College of Pharmacy, Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China, [4]French Associates Institute for Agriculture and Biotechnology of Drylands, Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Sede Boqer, Israel, [5]Biology Department, Brookhaven National Laboratory, Upton, NY, United States, [6]Centre for Modern Interdisciplinary Technologies, Nicolaus Copernicus University in Toruń, Toruń, Poland

KEYWORDS

metabolic engineering, synthetic biology, plants and algae, natural products, secondary metabolites, genetic modification, breeding

Editorial on the Research Topic
Metabolic engineering of valuable compounds in photosynthetic organisms

Photosynthetic organisms, including plants and algae, possess a remarkable ability to harness carbon dioxide and solar energy, enabling them to produce a vast array of complex compounds such as phenolic acids (Zhou et al., 2021), terpenes (Miller et al., 2020), unsaturated fatty acids (Kokabi et al., 2020; Gan et al., 2022), and other lipid products (Zienkiewicz and Zienkiewicz, 2020). This inherent capability positions them as highly promising platforms for the sustainable production of valuable biomolecules. While the industrial application of photosynthetic organisms in synthetic biology is not as advanced as that of model heterotrophs or mammalian systems, their significance as primary contributors to global biomass can be further developed. In fact, they are increasingly emerging as key players in the booming field of synthetic bioproducts, driven by advancements in genome editing tools and other innovative technologies. As we explore and exploit the potential of photosynthetic organisms, we open up exciting possibilities for the production of environmentally friendly and renewable biomaterials that can address pressing societal and ecological challenges.

This Research Topic includes eight original research and two review articles, with a special focus on the metabolic engineering of valuable biomaterials in plants and algae. Taparia et al., developed modular CRISPR/Cas9 constructs for the model diatom *Phaeodactylum tricornutum* that allow the multiplexed targeting and creation of marker-free genome-edited lines. The system was used to knock out *StLDP*, the gene encoding Stramenopile-type lipid droplet protein essential for lipid droplet biogenesis (Figure 1). Mellor et al. expressed human P450s in tobacco chloroplasts to produce indican, suggesting a strategy for producing high-value chemicals or drug metabolites in photosynthetic organisms (Figure 1). Another research ariticle investigated the biosynthesis of isoprenoids in poplar, and revealed that the 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR) and 1-deoxy-D-

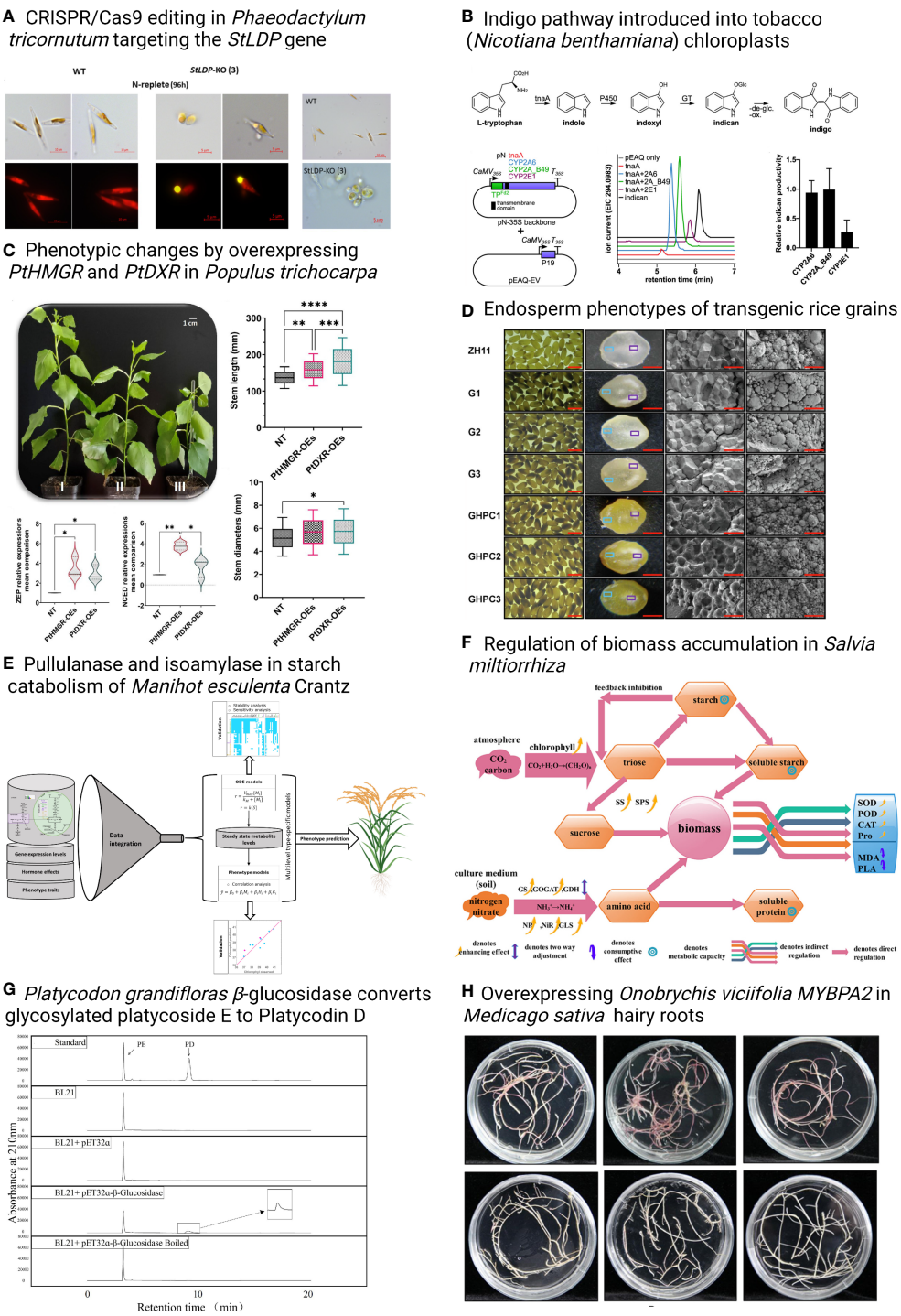


FIGURE 1
Overview of the original research articles in this Synthetic Biology Research Topic. Asterisks indicate statistically significant differences, *p <0.05, **p <0.01, ***p <0.001, ****p <0.0001

xylulose5-phosphate reductoisomerase (DXR) play important roles in regulating the genes in methylerythritol phosphate (MEP) and mevalonic acid (MVA) pathways and isoprenoids made from the MEP and MVA pathways (Figure 1) (Movahedi et al.). Li et al. produced carotenoids in rice (*Oryza sativa*) endosperm by overexpressing rice GOLDEN2-LIKE (*OsGLK*) transcription factor and *OsGLK* with three other carotenogenic genes, *tHMG1* (truncated *Saccharomyces cerevisiae* 3-hydroxy-3-methylglutaryl-CoA reductase), *ZmPSY1* (*Zea mays* L. phytoene synthase), and *PaCrtI* (*Pantoea ananatis* phytoene desaturase), to improve the nutritional

composition of rice (Figure 1). Another research article in rice developed models by multilevel mathematical modeling using the data from rice lines with genome modification in MVA pathways, providing tools that can help prioritize metabolic engineering strategies for specific metabolic goals through exogenous pathways (Figure 1) (Basallo et al.). In perennial herbs, Wang et al. identified physiological/biochemical indicators, such as enzyme activities of glutamine synthetase (GS), glutamate synthase (GLS), glutamate dehydrogenase (GDH), peroxidase (POD), and catalase (CAT), were related to biomass accumulation in *Salvia miltiorrhiza* (Figure 1); Su et al. characterized the β-glucosidase in *Platycodon grandifloras*, which can convert glycosylated platycoside E to Platycodin D *in vitro* (Figure 1); Jin et al. identified an MYB transcription factor OvMYBPA2 in *Onobrychis viciifolia* by transcriptome analyses and confirmed its function in the regulation of proanthocyanidins in transgenic *Medicago sativa* (Figure 1). Strand and Walker reviewed bioengineering from an energetics perspective using photosynthetic organisms for bioproducts of interest (Figure 1). Another review article discussed the recent progress in engineering fatty acids and storage lipids in various plant species and tissues and summarized an inventory of specific lipogenic factors for plant lipid products (Figure 1) (Cai et al.).

Author contributions

Z-YD: Funding acquisition, Resources, Visualization, Writing – original draft, Writing – review & editing. WB: Writing – review & editing. GK: Writing – review & editing. IK-G: Writing – review & editing. X-HY: Writing – review & editing. AZ: Writing – review & editing. KZ: Writing – review & editing.

Funding

Research of the Topic Editors is supported by the funds from NSF 2121410 (-YD), the DOE Office of Science, Office of BER, DE-SC0021369 (X-HY).

Acknowledgments

The editors would like to thank all reviewers who evaluated manuscripts and contributors to this Research Topic.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision

Publisher's note

References

Gan, L., Park, K., Chai, J., Updike, E. M., Kim, H., Voshall, A., et al. (2022). Divergent evolution of extreme production of variant plant monounsaturated fatty acids. *Proc. Natl. Acad. Sci. U.S.A.* 119, e2201160119. doi: 10.1073/pnas.2201160119

Kokabi, K., Gorelova, O., Zorin, B., Didi-Cohen, S., Itkin, M., Malitsky, S., et al. (2020). Lipidome remodeling and autophagic respose in the arachidonic-acid-rich microalga lobosphaera incisa under nitrogen and phosphorous deprivation. *Front. Plant Sci.* 11, 614846. doi: 10.3389/fpls.2020.614846

Miller, G. P., Bhat, W. W., Lanier, E. R., Johnson, S. R., Mathieu, D. T., and Hamberger, B. (2020). The biosynthesis of the anti-microbial diterpenoid leubethanol in *Leucophyllum frutescens* proceeds via an all- *cis* prenyl intermediate. *Plant J.* 104, 693–705. doi: 10.1111/tpj.14957

Zhou, W., Li, S., Maoz, I., Wang, Q., Xu, M., Feng, Y., et al. (2021). SmJRB1 positively regulates the accumulation of phenolic acid in Salvia miltiorrhiza. *Ind. Crops Prod* 164, 113417. doi: 10.1016/j.indcrop.2021.113417

Zienkiewicz, K., and Zienkiewicz, A. (2020). Degradation of lipid droplets in plants and algae—Right time, many paths, one goal. *Front. Plant Sci* 11. doi: 10.3389/FPLS.2020.579019

frontiers in Plant Science

ORIGINAL RESEARCH
published: 04 January 2022
doi: 10.3389/fpls.2021.793574

A Mathematical Model for Characterizing the Biomass and the Physiological/Biochemical Indicators of *Salvia miltiorrhiza* Based on Growth-Defense Tradeoff

Ke Wang[1,2], *Zhu-Yun Yan*[2]*, *Yuntong Ma*[2], *Bo Li*[2], *Wei Wang*[1], *Luming Qi*[3], *Hongmei Jia*[2], *Na Li*[2] and *Zhun Wang*[4]

[1] School of Big Data and Artificial Intelligence, Chengdu Technological University, Chengdu, China, [2] School of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, China, [3] School of Rehabilitation and Health Preservation, Chengdu University of Traditional Chinese Medicine, Chengdu, China, [4] TCM Health Industrial Technology Institute of Traditional Chinese Medicine, Chengdu, China

OPEN ACCESS

Edited by:
Guoyin Kai,
Zhejiang Chinese Medical University, China

Reviewed by:
Zhihua Liao,
Southwest University, China
Chenliang Yu,
Zhejiang Agriculture and Forestry University, China

***Correspondence:**
Zhu-Yun Yan
yanzhuyun@cdutcm.edu.cn

Specialty section:
This article was submitted to Plant Systems and Synthetic Biology, a section of the journal Frontiers in Plant Science

Received: *12 October 2021*
Accepted: *30 November 2021*
Published: *04 January 2022*

Citation:
Wang K, Yan Z-Y, Ma Y, Li B, Wang W, Qi L, Jia H, Li N and Wang Z (2022) A Mathematical Model for Characterizing the Biomass and the Physiological/Biochemical Indicators of Salvia miltiorrhiza Based on Growth-Defense Tradeoff. Front. Plant Sci. 12:793574. doi: 10.3389/fpls.2021.793574

Carbon(C) and nitrogen(N) metabolisms are important for plant growth and defense, and enzymes play a major role in these two metabolisms. Current studies show that the enzymes of N Metabolism, C Metabolism, and defense are correlated with biomass. Then, we conducted this research under the assumption that enzymes could characterize the relationship based on growth-defense tradeoff, and some of the enzymes could be used to represent the plant growth. From the mechanism model, we picked out 18 physiological/biochemical indicators and obtained the data from 24 tissue culture seedlings of *Salvia miltiorrhiza* (*S.miltiorrhiza*) which were grafted with 11 endophytic fungi. Then, the relationship between the biomass and the physiological/biochemical indicators was investigated by using statistical analysis, such as correlation analysis, variable screening, and regression analysis. The results showed that many physiological/biochemical indicators, especially enzyme activities, were related to biomass accumulation. Through a rigorous logical reasoning process, we established a mathematical model of the biomass and 6 key physiological/biochemical indicators, including glutamine synthetase (GS), glutamate synthase (GLS), glutamate dehydrogenase (GDH), peroxidase (POD), catalase (CAT), and soluble protein from Cobb-Douglas production function. This model had high prediction accuracy, and it could simplify the measurement of biomass. During the artificial cultivation of *S.miltiorrhiza*, we can monitor the biomass accumulation by scaling the key physiological/biochemical indicators in the leaves. Interestingly, the coefficients of Lasso regression during our analysis were consistent with the mechanism of growth-defense tradeoff. Perhaps, the key physiological/biochemical indicators obtained in the statistical analysis are related to the indicators affecting biomass accumulation in practice.

Keywords: enzyme activities, biomass, *Salvia miltiorrhiza*, mathematical model, growth-defense tradeoff

1. INTRODUCTION

As is known, Carbon (C) and nitrogen(N) are the main elements for the growth and defense of plants. Plants synthesize propane on the one hand through photosynthesis, which leads to starch, sucrose, and soluble sugars, and on the other hand they translate inorganic N into amino acids from ammonification, which leads to soluble proteins and are used for growth and defense in a certain C:N ratio (Lea and Morot-Gaudry, 2001). At the same time, plants consume the assimilated products of C/N metabolism through defense under biotic and abiotic stresses. Therefore, plants maintain a dynamic tradeoff between growth and defense by optimizing resource allocation through C/N metabolism and defense to enable plants to survive (Li et al., 2019).

Current studies show that some key enzyme activities of C/N metabolism have an available correlation with biomass accumulation (Noor et al., 2010). The activities of defensive enzymes show a significant correlation with adaptation to environmental stresses (Haddidi et al., 2020). Thus, the activities of sucrose phosphate synthase (SPS) and sucrose synthase (SS) in leaves, as well as the content of chlorophyll, soluble sugars, and starch are often used as important indicators to characterize the status of the C metabolism (Moriguchi et al., 1992; Verma et al., 2011). In N metabolism, nitrate reductase (NR) and nitrite reductase (NiR) translate N into NH_4^+, formulate glutamate and glutamine through ammonia assimilation (Liang et al., 2018; Salehin et al., 2019). Glutaminase (GLS), glutamine synthetase (GS), glutamate synthase (GOGAT), and glutamate dehydrogenase (GDH) can promote amino acid accumulation. But with a large increase in the reduction of NH_4^+ from NO_3^-, a sustained raising in GDH amination activity will inhibit GLS, GS, and GOGAT activities. Thus, GLS, GS, and GOGAT have a positive influence on biomass accumulation, while GDH has a bidirectional effect on biomass accumulation (Liang et al., 2018; Salehin et al., 2019; Gonzíalez-Moro et al., 2021; Wei et al., 2021). Many soluble proteins are crucial components of enzymes. They are involved in the physiological/biochemical metabolic and are the key indicators of whether plants suffered from heavy metal stress.

In the competition of biotic and abiotic stresses, the activities of enzymes, such as GS, GDH, NR, GOGAT, NiR, and GLS, as well as the soluble protein content, are considerable indicators of the N metabolic (Salehin et al., 2019; Gonzíalez-Moro et al., 2021; Wei et al., 2021). The enzymatic activities of superoxide dismutase (SOD), peroxidase (POD), catalase(CAT), proline (Pro), phenylalanine ammonialyase (PAL), and malondialdehyde (MDA) are commonly used to characterize defense responses (Jaafar et al., 2012; Farooq et al., 2020; Sarker and Oba, 2020; Zaheer et al., 2020). The increase of the enzymes activities such as SOD, POD, CAT, and Pro can significantly improve plant growth, biomass, chlorophyll content, and gas exchange properties (Sarker and Oba, 2020; Zaheer et al., 2020), while PAL and MDA inhibit plant growth by reducing the activity of antioxidant enzymes through oxidative stress (Jaafar et al., 2012; Farooq et al., 2020). Although the relationship between C/N metabolism and plant growth defense is obvious, there is still a lack of systematic research.

Based on the principles of metabolism, in this study we hypothesized that the correlation among the enzymes of C/N metabolic and defense could characterize the growth-defense tradeoff, and some of the enzymes could indicate the connection between the biomass and the physiological/biochemical indicators. Then, a mechanism model was established and 18 physiological/biochemical indicators were picked out (**Figure 1**). Since *S.miltiorrhiza* is considerable for the treatment of coronary heart disease and cerebrovascular disease (Su et al., 2015; Ma et al., 2016; Wang et al., 2017; Li, 2018; Shi et al., 2019), this study has practical significance to guide the production of medicinal plants. We cultured 24 tissue culture seedlings of *Salvia miltiorrhiza* (*S.miltiorrhiza*) which were grafted with 11 endophytic fungi. The 11 fungi are non-pathogenic and can intervene in physiological metabolism so that we can obtain data on different growth states. Through lasso screening variable (Efron et al., 2004) and regression analysis of plant physiological and biochemical indexes and biomass, the functional relationship between *S.miltiorrhiza* biomass and physiological /biochemical indexes was found. Thus, through the monitoring of relevant indicators of *S.miltiorrhiza*, we can understand the growth status of *S.miltiorrhiza*.

2. MATERIALS AND METHODS

2.1. Plant Materials

In this study, 24 tissue culture seedlings were inoculated from 11 strains of non-pathogenic endophytic fungi from 8 species of *S.miltiorrhiza* which our group obtained in previous studies (Ya-Li, 2018). The culture of tissue culture seedlings and endophytic fungal inoculation were obtained by the method in Lan et al. (2016) (daytime 25^{o}C/nighttime 20^{o}C, 14/10 h, light intensity 3,000 LX) where young leaves of *S.miltiorrhiza* were employed as explants. The seedlings with 3–5 roots were transplanted into artificial soil containing 50 ml of 1/2 MS medium, sealed with a sealing film, and incubated in vertical light for 3 days (Hesheng, 2000), then a small piece of fungal cake was taken with a hole punch and placed on the root of the seedlings in the culture flask, and the root of *S.miltiorrhiza* was gently punctured with a sterile needle. After 30 days of incubation, the plants were removed and the biomass and related physiological indicators were measured.

2.2. Determination of the Biomass

The *S.miltiorrhiza* seedlings were taken out and washed carefully to remove impurities attached to the roots. After absorbing the surface moisture with absorbent paper, weigh it and subtract the weight of seedlings when transplanting to get the net biomass.

Abbreviations: C, carbon; CAT, catalase; GDH, glutamate dehydrogenase; GLS, glutaminase; GOGAT, glutamate synthase; GS, glutamine synthetase; MDA, malondialdehyde; N, nitrogen; NiR, nitrite reductase; NR, nitrate reductase; PAL, phenylalanine ammonialyase; PGA, 3- phosphoglyceric acid; POD, peroxidases; Pro, proline; NADPH, reducing agent; *S.miltiorrhiza*, *Salvia miltiorrhiza*; SOD, superoxide dismutase; SPS, synthase; SS, sucrose synthase.

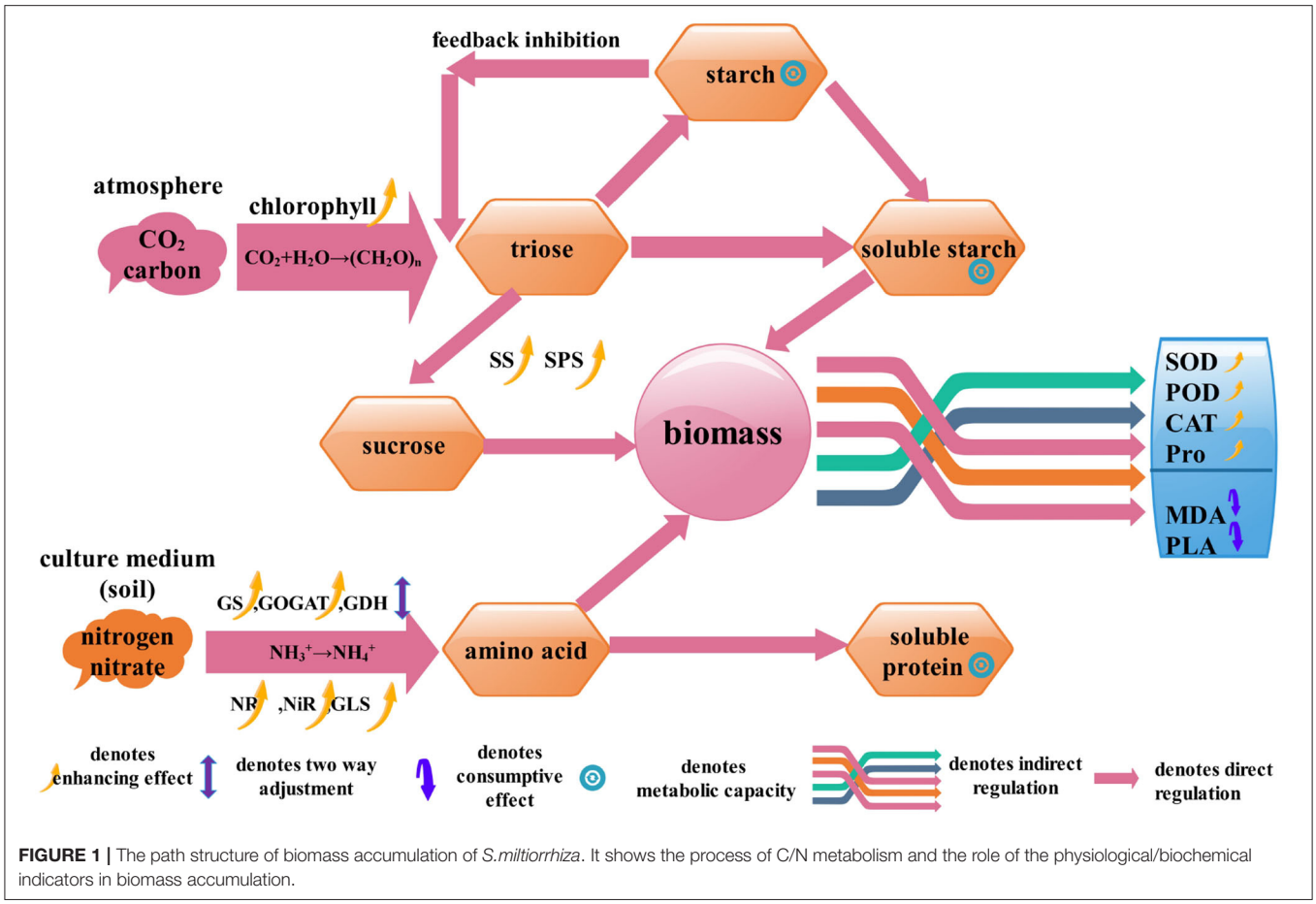


FIGURE 1 | The path structure of biomass accumulation of *S.miltiorrhiza*. It shows the process of C/N metabolism and the role of the physiological/biochemical indicators in biomass accumulation.

2.3. Determination of the Physiological/Biochemical Indicators

The activities of SS, SPS, NR, GS, and GOGAT were determined through enzyme solution prepared from young leaves of *S.miltiorrhiza* (Deane-Drummonda et al., 1979; Islam et al., 1996; Li, 2000; Zhao et al., 2003). The contents of reducing sugar and soluble sugar were estimated by adopting 3,5-dinitrosalicylic acid method and anthrone method, respectively (Zhang, 1990; Li, 2000; Gao, 2006). The soluble protein content was assessed with the coomassie brilliant blue method (Elsharkawy et al., 2012). Chlorophyll content was determined by spectrophotometry (Strain and Svec, 1966). The activities of SOD, POD, CAT, PAL, and MDA and Pro in *S.miltiorrhiza* leaves were measured according to Zhu et al. (1983), Polle et al. (1994), and Gao (2006).

2.4. Statistical Analysis

2.4.1. Correlation Analysis

Because of the large difference in unit and quantity, the physiological/biochemical indicators were standardized by the following function:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad (1)$$

where x_{ij} was the value of the jth physiological/biochemical indicator of the ith plant of *S.miltiorrhiza*, $\bar{x}_j$ denoted the average of jth indicator, and s_j denoted the SD of the jth indicator.

The correlation analysis between the physiological/biochemical indicators of *S.miltiorrhiza* and the biomass included linear correlation analysis and nonlinear correlation analysis, so the correlation coefficient matrix analysis was chosen to be used. It included three aspects: (1) analysis of the variation between the biomass and the physiological/biochemical indicators; (2) linear correlation between the biomass and the physiological/biochemical indicators; and (3) analysis of the nonlinear relationship between the biomass and the physiological/biochemical indicators.

2.4.2. Systematic Analysis Based on Lasso Algorithm

According to the mechanism analysis, we tried to find the physiological/biochemical indicators related to the biomass accumulation of *S.miltiorrhiza*, so there would be duplication or the introduction of irrelevant factors. However, there are multiple covariates among various physiological/biochemical indicators, in which case least squares and partial least squares work poorly. In order to select a concise set of physiological/biochemical indicators to effectively predict biomass and improve the prediction accuracy of the model (Efron et al., 2004), Lasso

regression was introduced for variable screening. The regression function was as follow:

$$y_i = \sum_{j=1}^{p} \beta_j x'_{ij} + \varepsilon_i, (i = 1, 2, ..., n), \quad (2)$$

where x'_{ij} denoted the jth physiological/biochemical indicator of the ith sample, y_i denoted the biomass of *S.miltiorrhiza* of the ith sample, and p denoted the number of physiological/biochemical indicators screened.

Because $X_i = (x_{i1}, \ldots, x_{ip})^T, (i = 1, 2, ..., p)$ were multicollinearity, the Lasso method was used to filter variables. The Lasso constructed a penalty function to obtain a more refined model, which made it compress some regression coefficients. Here, l_1-penalty function was used for regularization estimation parameter as in Efron et al. (2004), defined as

$$\hat{\beta}(\lambda) = \text{argmin} \frac{1}{N} ||Y - X\beta||^2 + \lambda \sum_j |\beta_j|, \quad (3)$$

where N denoted the number of samples, $Y = (y_1, y_2, ..., y_N)^T$ denoted the biomass, $X = (X_1, X_2, ..., X_N)^T$ denoted the physiological/biochemical indicators, β denoted the regression coefficient, and $\lambda \geq 0$ denoted the penalty parameter.

2.4.3. Nonlinear Regression Analysis Based on Cobb-Douglas Production Function

Based on the characteristics of important variables screened by Lasso, nonlinear regression analysis was utilized since correlations and interactions between variables still existed. Assuming that y denotes biomass, $A_i (i = 1, 2, \cdots n)$ denotes factors with positive influence coefficients, and $\beta_j (j = 1, 2, \cdots m)$ denotes factors with negative influence coefficients, similar to the Cobb-Douglas production model in Carter (2012), the models for the biomass and the physiological/biochemical indicators were established as follows:

$$Y = H \cdot \frac{A_1^{a_1} A_2^{a_2} \cdots A_n^{a_n}}{B_1^{b_1} B_2^{b_2} \cdots B_n^{b_m}}, \quad (4)$$

Where $H, a_1, a_2, \cdots, a_n, b_1, b_2, \cdots, b_m$ denoted parameters to be determined by the Equation (4). Then took common logarithms on both sides of the Equation (4) at the same time, that was

$$\ln Y = \ln H + a_1 \ln A_1 + a_2 \ln A_2 + \cdots + a_n \ln A_n - b_1 \ln B_1 - b_2 \ln B_2 - \cdots - b_m \ln B_m. \quad (5)$$

The Equation (4) was transformed into a linear regression, from which the initial values of the nonlinear regression parameters could be obtained from the Equation (5). Then, the solution of Equation (4) could be optimized by performing a nonlinear regression.

2.4.4. Data Analysis Environment

All the data were analyzed in R4.0.3 + Rstudio (Chang, 2013; Lantz, 2013). glmnet() was used for filtering variables in Lasso, ls()/nls() was used for linear/nonlinear regression analysis, and ggplot() was used for graph plotting.

3. RESULTS

3.1. Experimental Results and Data

To analyze the relationship between the biomass and the physiological/biochemical indicators, 24 samples of *S.miltiorrhiza* were collected according to the methods described in 2.1 ∼ 2.3, and the data of the 18 physiological/biochemical indicators such as chlorophyll, SS, SPS, soluble sugar, and the values of corresponding biomass were displayed in **Table 1**.

3.2. Descriptive Statistical Analysis of the Biomass and the Physiological/Biochemical Indicators

We defined biomass as the dependent variable(y) and chlorophyll(x_1), SS(x_2), SPS(x_3), soluble sugar(x_4), starch(x_5), GOGAT(x_6), GS(x_7), NR(x_8), NiR(x_9), GLS(x_{10}), GDH(x_{11}), soluble protein(x_{12}), POD(x_{13}), SOD(x_{14}), CAT(x_{15}), Pro(x_{16}), PAL(x_{17}), and MDA(x_{18}) as independent variables. To eliminate the difference in magnitude between variables, the criteria were standardized by using Equation (1) and then other corresponding analyses were performed.

3.2.1. Differences in Some Physiological/Biochemical Indicators and Lack of Consensus in the Biomass Accumulation

Some physiological/biochemical indicators were influenced by individual plants, especially SS(x_2) and GOGAT(x_6) were very different (**Figure 2**). It indicated that they lacked consensus in the biomass accumulation of *S.miltiorrhiza* and there was uncertainty among indicators.

3.2.2. Many Physiological/Biochemical Indicators Have a Linear (or Nonlinear) Correlation With the Biomass

From the result of correlation analysis, we can see there was a positive correlation of the biomass(y) with chlorophyll(x_1) and CAT(x_{15}) (**Figure 3**), and a negative correlation with GLS(x_{10}) and POD(x_{13}). Considering the relationship between the physiological/biochemical indicators, chlorophyll(x_1), NR(x_8), and CAT(x_{15}) were positively correlated, while chlorophyll(x_1) was negatively correlated with SPS(x_3), NiR(x_9), and POD(x_{13}). SPS(x_3) was positively correlated with CAT(x_{15}), while SPS(x_3) was negatively correlated with NiR(x_9), GLS(x_{10}) and POD(x_{13}). Similarly, these results indicated that linear correlations existed between the biomass and the physiological/biochemical indicators.

There was a certain nonlinear relationship (**Figure 4**) between the biomass and the physiological/biochemical indicators. For instance, there was a nonlinear relationship between the biomass(y), and NiR(x_9), and soluble protein(x_{12}). Thereby, biomass could not be expressed by a single enzyme activity or

TABLE 1 | Biomass and physiological/biochemical indicators of *S. miltiorrhiza*.

No	Biomass	Chlorophyll	SS	SPS	Soluble sugar	Starch	GOGAT	GS	NR	NIR	GLS	GDH	Soluble	POD	SOD	CAT	PRO	PAL	MDA
1	1.04	187.38	100.24	94.18	173.60	231.12	48.89	149.78	26.73	62.88	141.30	46.08	10.79	102.40	21.51	553.86	82.61	100.92	227.21
2	1.41	152.56	58.98	110.35	82.81	227.04	56.84	72.18	20.70	83.66	70.04	34.29	7.37	103.87	16.21	403.92	45.96	93.84	238.91
3	1.39	202.88	47.25	120.05	63.48	192.40	296.44	258.31	18.37	44.65	86.70	73.73	13.05	79.33	18.25	504.90	60.10	169.62	263.68
4	0.98	160.17	82.81	102.36	158.63	196.75	79.82	112.13	39.66	62.61	72.69	66.87	13.42	203.47	26.87	621.18	91.93	152.38	204.85
5	2.03	174.68	88.90	118.91	132.40	289.19	70.97	152.65	28.93	80.51	81.09	100.09	7.45	128.40	19.29	605.88	67.29	203.37	256.28
6	1.32	214.34	40.92	139.27	148.14	249.62	102.82	71.57	37.03	85.81	102.21	61.51	8.31	110.80	19.83	563.04	88.23	103.19	235.12
7	2.25	160.84	78.52	148.21	86.62	221.64	91.91	178.59	19.16	58.31	110.54	36.22	10.51	88.80	20.83	615.06	40.77	134.70	176.13
8	1.19	166.41	100.29	152.40	161.93	182.49	105.90	71.67	34.65	68.46	121.30	26.58	10.68	98.67	25.88	553.86	61.72	172.85	257.14
9	1.23	203.04	170.71	130.52	66.56	239.29	64.09	50.11	16.64	88.85	112.71	16.93	8.62	96.67	17.38	584.46	43.71	131.07	184.38
10	1.07	214.27	33.39	145.55	140.28	233.01	43.08	58.01	35.07	65.61	144.62	38.15	11.88	127.87	24.34	434.52	68.36	208.06	240.46
11	0.94	191.76	88.23	87.52	116.16	238.10	29.62	48.53	29.04	90.13	140.57	13.93	8.29	114.40	16.33	511.02	63.05	135.70	232.54
12	1.25	160.38	22.77	94.75	57.85	198.83	35.11	39.05	14.46	62.62	101.60	24.65	12.20	96.40	26.20	593.64	64.38	121.54	310.63
13	0.48	116.58	90.26	355.59	162.98	247.04	125.73	101.31	20.37	90.47	194.85	20.58	8.64	248.11	18.44	241.74	53.38	105.26	221.54
14	1.06	103.47	120.38	770.74	203.46	333.85	201.82	82.14	25.43	105.50	215.34	42.65	8.00	246.11	25.22	91.80	52.60	124.82	248.37
15	0.86	130.59	155.42	614.54	128.79	257.06	212.45	78.38	16.10	87.80	253.22	75.66	8.84	252.37	19.20	192.78	67.72	157.39	328.52
16	0.54	89.78	88.48	882.80	98.80	249.39	39.91	66.06	12.35	89.51	173.20	19.29	8.84	169.71	23.68	250.92	36.53	153.42	181.46
17	0.65	110.24	93.60	977.17	82.86	237.87	35.75	60.58	10.36	99.08	174.48	52.08	7.73	168.64	22.43	293.76	34.67	145.00	301.86
18	0.65	132.35	80.32	243.15	96.08	277.75	83.20	52.02	12.01	89.12	300.71	44.37	11.91	163.84	25.88	250.92	47.79	225.36	312.70
19	0.86	98.60	80.11	502.85	72.80	261.91	83.31	128.01	9.10	89.34	208.74	28.08	11.85	167.17	20.42	302.94	57.02	94.13	202.96
20	0.77	92.34	131.02	203.58	134.67	275.86	64.34	117.74	16.83	93.30	148.66	29.58	9.39	166.24	20.02	149.94	61.75	117.86	193.50
21	0.66	127.12	101.46	248.29	74.47	268.69	72.29	70.85	9.31	70.70	80.47	31.29	10.11	164.91	18.39	272.34	43.36	168.17	347.96
22	0.98	152.86	135.44	567.16	151.44	305.34	270.98	109.18	18.93	90.69	145.72	95.38	9.61	170.24	36.83	122.40	77.40	140.63	236.84
23	1.19	140.67	160.92	305.18	212.24	301.72	306.41	163.26	26.53	93.62	151.27	65.37	7.85	166.51	37.67	198.90	64.15	110.90	170.11
24	0.81	82.39	94.16	192.73	90.83	237.21	51.44	62.29	11.35	101.33	105.87	29.36	7.73	167.04	22.45	281.52	49.63	133.27	408.67

In this table biomass, chlorophyll, SS, SPS, soluble sugar, starch, glutamate synthase (GOGAT), glutamine synthetase (GS), nitrate reductase (NR), nitrite reductase (NIR), glutaminase (GLS), glutamate dehydrogenase, soluble, peroxidases (POD), superoxide dismutase (SOD), catalase (CAT), proline (PRO), phenylalanine ammonialyase (PAL), malondialdehyde (MDA) are separately measured as (g), (mg· g^{-1} · FW), (mol· min^{-1} · g^{-1} · FW), (mol· min^{-1} · g^{-1} · FW), (U· g^{-1} · FW), (g· g^{-1} · FW), (U· g^{-1} · FW), (nmol· min^{-1} · g^{-1} · FW), (mol· h^{-1} · g^{-1} · FW), (mol· min^{-1} · g^{-1} · FW), (mol· h^{-1} · g^{-1} · FW), (nmol· min^{-1} · g^{-1} · FW), protein(g· ml^{-1}), (U· g^{-1} · h· FW), (U· g^{-1} · h· FW), (nmol· min^{-1} · g^{-1} · FW), (U· g^{-1} · FW), (U· mg^{-1} · h), and (nmol· g^{-1} · FW), respectively.

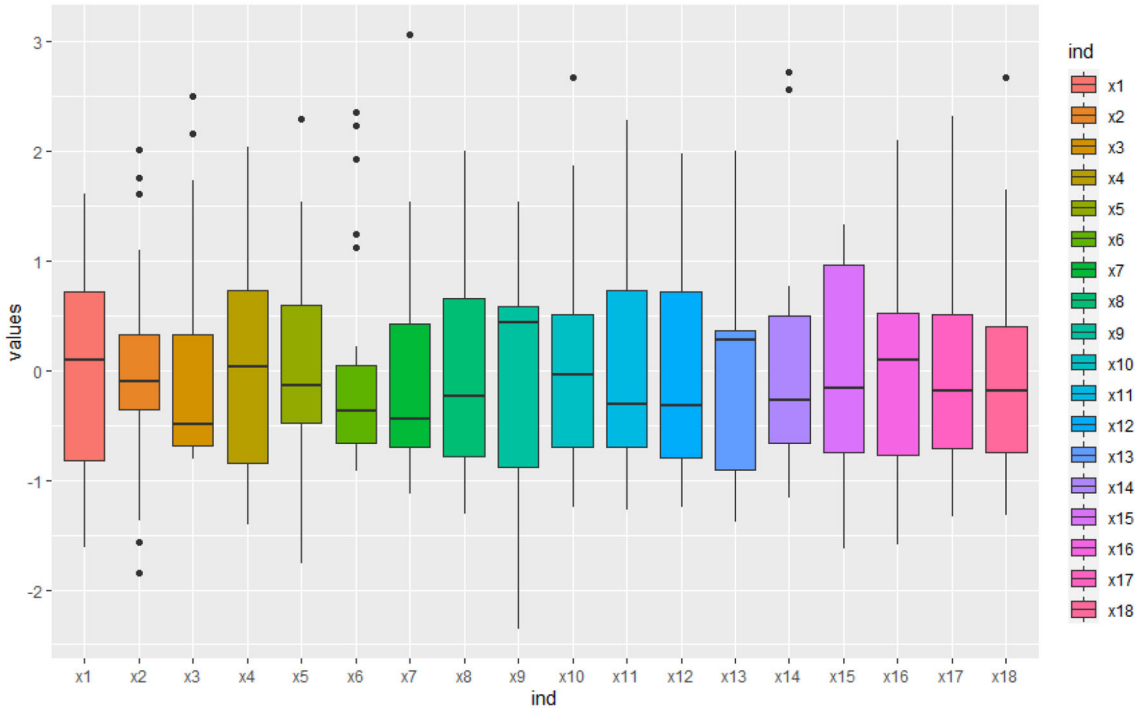


FIGURE 2 | The box line diagram of the physiological/biochemical indicators. The blue curve is the regression curve, the shaded part indicates the confidence interval, and the larger the shaded area the worse the regression effect of the curve. It shows that some physiological/biochemical indicators lack consensus in the biomass accumulation of *S.miltiorrhiza*.

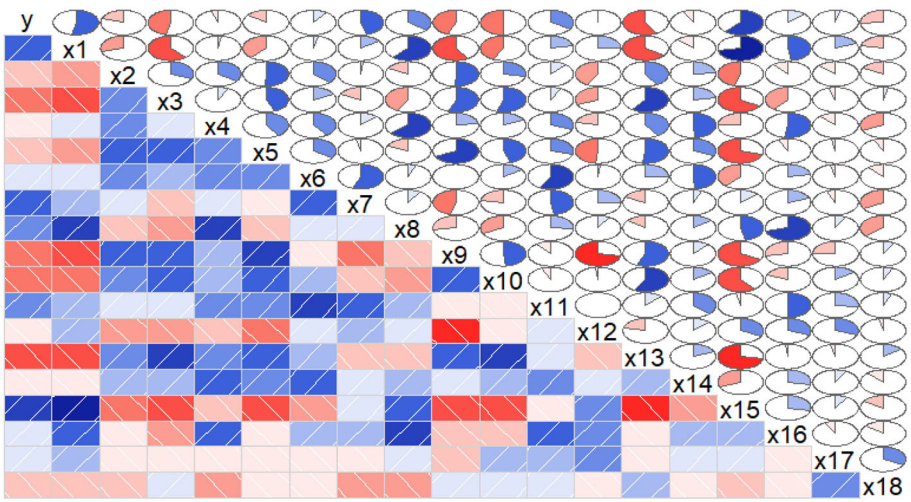


FIGURE 3 | Correlations between the biomass and the physiological/biochemical indicators of *S.miltiorrhiza* biomass. The blue rectangles represent a positive correlation, the red rectangles represent a negative correlation. The area of the circle represents the degree of correlation. According to correlograms shown in this figure, there were correlations among the biomass and the physiological/biochemical indicators.

physiological/biochemical indicator, but by a combination of some effective physiological/biochemical indicators (**Figures 3**, **4**). Therefore, it was necessary to systematically analyze the relationship between the physiological/biochemical indicators and the biomass.

3.3. Some Physiological/Biochemical Indicators Play a Major Role in the Biomass Accumulation

We used the glmnet() of the R software to calculate the Lasso model (parameters set to default values) and optimized the model

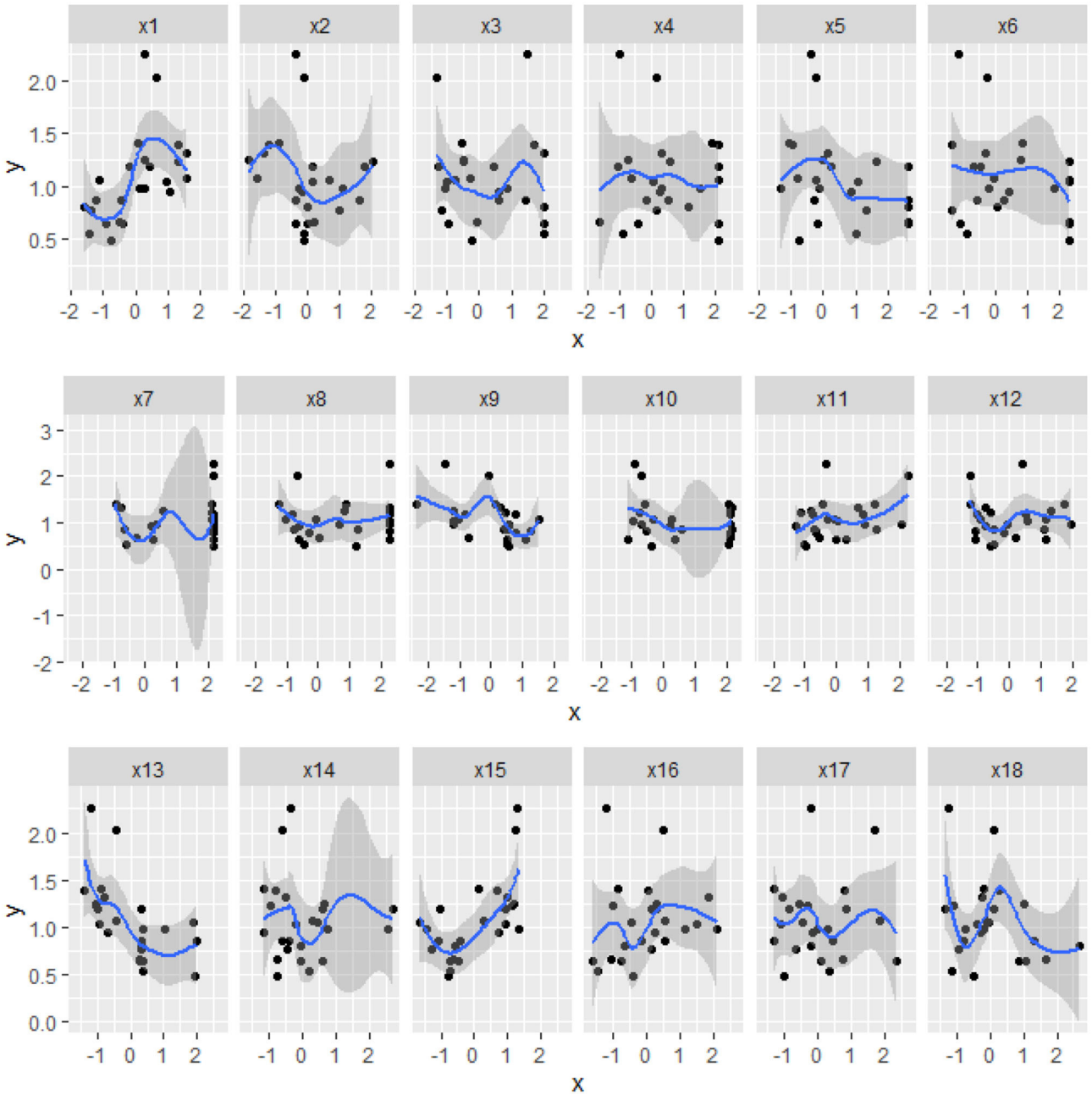


FIGURE 4 | The scatter and regression curves of the biomass and the physiological/biochemical indicators of *S.miltiorrhiza*. The horizontal axis of x_i, is the physiological/biochemical indicator of i and the vertical axis is y, which is the biomass of *S.miltiorrhiza*. The blue curve is the regression curve, the shaded part indicates the confidence interval, and the larger the shaded area indicates the worse regression effect of the curve.

and function by cv.glmnet() to obtain the results shown in **Figure 5** and **Table 2**.

The results showed that GS(x_7), GDH(x_{11}), and CAT(x_{15}) were positively correlated with biomass and acted as promoters when the whole system was considered. GLS(x_{10}), POD(x_{15}), and soluble protein(x_{13}) were negatively correlated with biomass accumulation and acted as inhibitors.

3.4. Some Key Physiological/Biochemical Indicators Are Able to Express the Biomass Well

A nonlinear regression analysis was performed by using the Douglas production model with the influencing factors of GS(x_{15}), GLS(x_{15}), GDH(x_{15}), soluble protein(x_{15}), POD(x_{15}), and CAT(x_{15}), and biomass. Took 5/6 of the data as the training set and 1/6 as the verification set, repeated the training 1,000

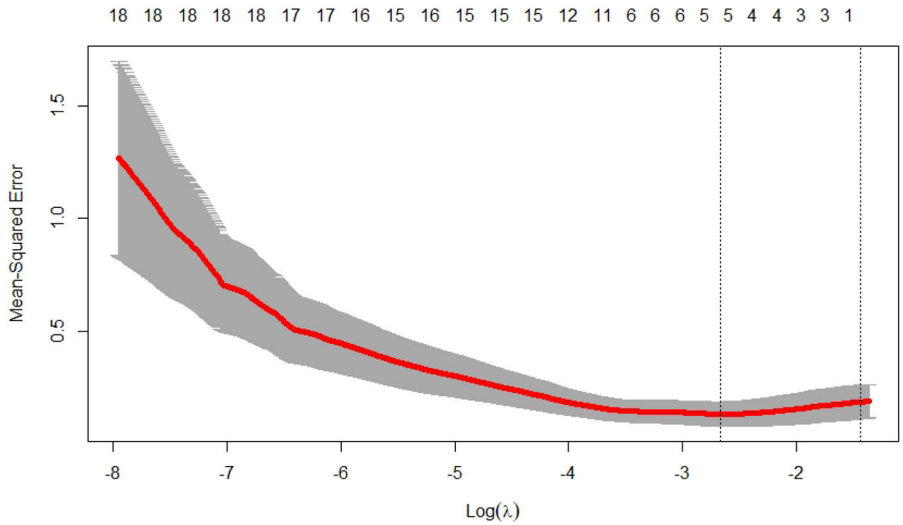


FIGURE 5 | Lasso regression analysis number of iterations vs. mean square error. It shows that the optimal λ value is obtained in 5th step.

times, and selected the regression equation with the smallest verification set error. The model was as follows:

$$Y = 5.05853 \times \frac{x_7^{0.24737} x_{11}^{0.09026} x_{15}^{0.22631}}{x_{10}^{0.14776} x_{12}^{0.64021} x_{13}^{0.44381}}. \quad (6)$$

The predicted values of biomass could be obtained from Equation (6) and the results were shown in **Table 2**. The test showed that the prediction effect of the model was good (**Tables 3**, **4**). It showed that the regression equation was relatively successful and was able to predict the corresponding the biomass from the 6 important physiological/biochemical indicators.

4. DISCUSSION

Carbon/nitrogen metabolism are critical in growth and defense (Li et al., 2019). Adequate management of the plant in C/N metabolism and growth defense have a significant impact on crop productivity. In this study, we constructed a growth mechanism model by *S.miltiorrhiza* and picked out the relationship between the 18 physiological/biochemical indicators and the C/N metabolism, with the aim of finding important metabolic traits as an indication of the biomass (**Table 2**).

The variability of the physiological/biochemical indicators was analyzed by box-line plots (**Figure 2**), and it was observed that some indicators showed no consensus. That was, some physiological/biochemical indicators were greatly affected by individuals. In particular, SS(x_2) and GOGAT(x_6) were more discrete. SS(x_2) was an important enzyme catalyzing sucrose synthesis in plants, and sucrose was broken down into soluble sugars for drought resistance in plants on the one hand, and synthesized biomass with soluble proteins on the other hand, making SS showed uncertainty on the biomass (Kaur et al., 2006; Zhang et al., 2018). Similarly, GOGAT(x_6) consuming amino acids used by plants for the biomass accumulation during growth defense, formed aspartic acid, which was used for plant drought resistance under the action of Pro and producing glutamate, soluble protein, which leads to biomass accumulation. The uncertainty of the two conversion pathways, SS(x_2), and GOGAT(x_6), prevents them from being used as important indicators for characterizing the biomass (Wei et al., 2021).

There were linear(or nonlinear) correlations between the biomass and the physiological/biochemical indicators (**Figures 2**, **3**), indicating interdependence in the growth-defense tradeoff of plants. For example, there was a positive correlation between biomass(y) and chlorophyll(x_1)/CAT(x_{15}), indicating that chlorophyll was the main substance for photosynthesis in plants, and it played an important role in C metabolism (Luo, 2018). Although CAT(x_{15}) was mainly used in the defense, it reduced the toxicity of H_2O_2 (Abogadallah, 2010; Sarker and Oba, 2020), and maintained normal growth and defense. The biomass was negatively correlated with some physiological/biochemical indicators, such as GLS(x_6) promoting amino acid accumulation, and POD(x_{13}) depleting C and N assimilation products, affecting the biomass accumulation. POD(x_{13}) reduces H_2O_2 and converts carbohydrates to lignin, improving the physical defense barrier and depleting assimilates required for growth. Considering the relationship between the physiological/biochemical indicators, chlorophyll(x_1), NR(x_{15}), and CAT(x_{15}) were positively correlated, while chlorophyll(x_{15}) was negatively correlated with SPS(x_3), NiR(x_9), and POD(x_{13}) (**Figure 3**). On the one hand, chlorophyll is used by plants for photosynthesis to convert CO_2 in the air into C required by plants, and on the other hand, about 90% of N in plants, coming from biological N fixation, is reduced from nitrate to nitrite in soil by NR. Plant cells can rapidly transfer nitrite produced by nitrate reduction from the stroma to the chloroplasts of leaf cells or the plastids of root cells and are reduced in the chloroplasts by NiR to ammonia, while the potential toxicity of

TABLE 2 | Nonzero regression coefficient of Lasso.

x_7	x_{10}	x_{11}	x_{12}	x_{13}	x_{15}
1.385218e-03	−1.205714e-05	1.999661e-03	−4.228444e-03	−1.422853e-03	7.330554e-04

TABLE 3 | The results of the training set of the biomass and the key physiological/biochemical indicators.

x7	x10	x11	x12	x13	x15	y	hat_y	absolute_error
48.53	140.57	13.93	8.29	114.4	511.02	0.94	1.043	0.103
62.29	105.87	29.36	7.73	167.04	281.52	0.81	0.956	0.146
71.67	121.3	26.58	10.68	98.67	553.86	1.19	1.151	−0.039
82.14	215.34	42.65	8	246.11	91.8	1.06	0.609	−0.451
112.13	72.69	66.87	13.42	203.47	621.18	0.98	0.969	−0.011
258.31	86.7	73.73	13.05	79.33	504.9	1.39	1.728	0.338
101.31	194.85	20.58	8.64	248.11	241.74	0.48	0.72	0.24
117.74	148.66	29.58	9.39	166.24	149.94	0.77	0.817	0.047
71.57	102.21	61.51	8.31	110.8	563.04	1.32	1.425	0.105
178.59	110.54	36.22	10.51	88.8	615.06	2.25	1.63	−0.62
50.11	112.71	16.93	8.62	96.67	584.46	1.23	1.198	−0.032
152.65	81.09	100.09	7.45	128.4	605.88	2.03	1.898	−0.132
60.58	174.48	52.08	7.73	168.64	293.76	0.65	0.934	0.284
128.01	208.74	28.08	11.85	167.17	302.94	0.86	0.796	−0.064
72.18	70.04	34.29	7.37	103.87	403.92	1.41	1.476	0.066
66.06	173.2	19.29	8.84	169.71	250.92	0.54	0.771	0.231
39.05	101.6	24.65	12.2	96.4	593.64	1.25	0.952	−0.298
109.18	145.72	95.38	9.61	170.24	122.4	0.98	0.832	−0.148
149.78	141.3	46.08	10.79	102.4	553.86	1.04	1.387	0.347
78.38	253.22	75.66	8.84	252.37	192.78	0.86	0.68	−0.18

TABLE 4 | The results of the validation set of the biomass and the key physiological/biochemical indicators.

x7	x10	x11	x12	x13	x15	y	hat_y	absolute_error
58.01	144.62	38.15	11.88	127.87	434.52	1.07	0.866	−0.204
52.02	300.71	44.37	11.91	163.84	250.92	0.65	0.606	−0.044
70.85	80.47	31.29	10.11	164.91	272.34	0.66	0.869	0.209
163.26	151.27	65.37	7.85	166.51	198.9	1.19	1.134	−0.056

nitrite is broken down by CAT(x_{15}). Chlorophyll(x_1), NR(x_8), NiR(x_9), and CAT(x_{15}) collaborate with each other in C/N metabolism as well as in the growth-defenses tradeoff (Zhou et al., 2021). SPS(x_3) and SS(x_2) are both important enzymes for catalyzing sucrose synthesis in plants, however, SPS exhibiting different characteristics from sucrose SPS SS(x_2) was positively correlated with CAT(x_{15}), indicating that it promoted the biomass accumulation (**Figure 3**). While SPS was negatively correlated with NiR(x_9), GLS(x_{10}), and POD(x_{13})(**Figure 3**). It showed inhibition of biomass accumulation, but NiR(x_{15}) reduced nitrate in soil and then synthesized NH_4^+ for the biomass formation, and GLS(x_{10}) formed NH_4^+ into amino acids, which were against the previous analysis. All of these analyses indicated that biomass accumulation was a relatively complex process that required systematic analysis of the physiological/biochemical indicators affecting the biomass accumulation.

By Lasso analysis (**Figure 5** and **Table 2**), GS(x_7) and GDH(x_{11}) were positively correlated with biomass (**Table 3**), indicating that GS(x_7) and GDH(x_{11}) were the main substrate in the seedling stage facilitate protein synthesis with amino acids and were conducive to coordinate C metabolism (Kaur et al., 2006), which was consistent with the rapid assimilation of N in early growth and laid the foundation for high rate C assimilation in later growth stages (Salehin et al., 2019; Gonzíalez-Moro et al., 2021; Wei et al., 2021). GS(x_7) was responsible for primary NH_4^+ assimilation as well as germination, and elevating of GS(x_7) in leaves contributed to the conversion of inorganic NH_4^+ to organic N (Lea and Morot-Gaudry, 2001). There was a positive correlation between GDH(x_{11}) and total amino acid content in the root system (Gonzíalez-Moro et al., 2021). When GDH(x_{11}) was sufficiently induced in leaves, it would indicated substrate specificity in plants with other enzymes. Although

GLS facilitated amino acid accumulation, it was negatively correlated with the biomass (**Table 3**), possibly due to the lack of sufficient reducing agent(NADPH) or the unavailability of reducing agent(NADPH), making it possible that NH_3^+ might not attack the ketone group, leading to its release into the environment (Yoneyama et al., 2015). Soluble proteins were negatively correlated with biomass (**Table 3**), indicating that soluble proteins would consume amino acids used for biomass accumulation. POD was negatively correlated with biomass (**Table 3**), indicating that glutamate synthesis, as well as the synthesis of metabolic enzymes represented by POD, consumed N and C assimilation products, thus affecting the biomass accumulation. POD reduced H_2O_2 and converted carbohydrates to lignin, improving the physical defense barrier and depleting assimilates required for growth. However, despite its synthesis of N and C assimilation products, CAT(x_{15}) specifically reduced the toxicity of H_2O_2 (Abogadallah, 2010; Sarker and Oba, 2020), maintained cell membrane stability, and promoted growth and defense. Meanwhile, it was a positive correlation between CAT(x_{15}) and biomass(y) (**Table 3**).

The relationship between the physiological/biochemical indicators in the biomass accumulation mechanism model was consistent with the results of the Lasso model analysis. Our analysis not only illustrated that the biomass accumulation was related to some key physiological/biochemical indicators, but also the characteristics of these enzymes were the same as those of Lasso analysis. For example, GS(x_7) and GDH(x_{10}), which had positive coefficients in the Lasso model, were also present in the routes of C/N metabolism to promote biomass accumulation, and GLS(x_{11}) and POD(x_{13}), which had negative coefficients in the Lasso model, were also present in the routes of defense to consume the C and N and affect the biomass accumulation (Lea and Morot-Gaudry, 2001). Furthermore, although CAT(x_{15}) consumed C and N on the defense route, its negative synergistic effect with POD(x_{13}), or competition for substrates, could explain this phenomenon. Thereby, perhaps increasing GS(x_7), GDH(x_{10}), and CAT(x_{15}) activities, or decreasing GLS(x_{11}) and POD(x_{13}) activities, would be beneficial to promote biomass accumulation of *S.miltiorrhiza* .

Finally, we established a functional equation between the 6 key physiological/biochemical indicators and the biomass based on the Cobb-Douglas economic model. Although Cobb-Douglas economic model is an economic mathematical model used to predict the production of national and regional industrial systems or large enterprises and analyze the ways to develop production, this study skillfully uses it to establish the relationship between physiological/biochemical indicators and biomass. Interestingly, the correlation between the physiological/biochemical indicators and the biomass in this equation was the same as that obtained by Lasso regression. That is GS(x_7), GDH(x_{11}), and CAT(x_{15}) were positively correlated with the biomass, and GLS(x_{10}), POD(x_{12}), and soluble protein(x_{13}) were negatively correlated with biomass accumulation. The biomass predicted from the physiological/biochemical indicators of this equation had less error. In the artificial cultivation of *S.miltiorrhiza*, only the values of the key physiological/biochemical indicators of the above-ground parts are needed to obtain the corresponding biomass.

5. CONCLUSION

This study provides a rigorous logical reasoning process in terms of the selection of factors affecting biomass accumulation, the screening of key factors, and the establishment and validation of regression models. The method is applicable not only to metabolic engineering but also to phenomena with similar mechanistic features, such as the relationship between the activities and soil environment, and the self-organization of microbial communities. In addition, the generalized application of Lasso regression and the Cobb-Douglas production model used in this study provides a powerful tool for a comprehensive and systematic study of growth and active ingredient synthesis.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

ZY and KW obtained funding. BL, HJ, NL, and ZW designed the experiments and analyzed the data, with the physiology support from YM and LQ generated the structure diagram of biomass accumulation mechanism, WW modified Lasso and nonlinear regression algorithm. KW completed all the program design and graphics drawing. KW and ZY wrote this study with the contributions of all the authors, who also approved the final version prior to submission.

FUNDING

This study was supported by Sichuan Science and Technology Program (2021 YFS0045), Special Scientific and Technological Research Project of Sichuan Province Administration of Traditional Chinese Medicine (Project 2018QN013), State Key Laboratory of Characteristic Chinese Medicine Resources in Southwest China, School of Pharmacy, Chengdu University of Traditional Chinese Medicine, Chengdu, China.

REFERENCES

Abogadallah, G. (2010). Antioxidative defense under salt stress. *Plant Signal. Behav.* 5, 369–374. doi: 10.4161/psb.5.4.10873

Carter, S. (2012). "On the Cobb-Douglas and all that... ": the solow-simon correspondence over the aggregate neoclassical production function. *J. Post Keynesian Econ.* 34, 255–273. doi: 10.2753/PKE0160-3477340204

Chang, W. (2013). *R Graphics Cookbook*. Sebastopol: O'Reilly Media, Inc.

Deane-Drummonda, C. E., Clarksona, D. T., and Johnsonb, C. B. (1979). Effect of shoot removal and malate on the activity of nitrate reductase assayed *in vivo* in barley roots (Hordeum Vulgare cv. midas). *Plant Physiol.* 64, 660–662. doi: 10.1104/pp.64.4.660

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., Ishwaran, H., Knight, K., et al. (2004). Least angle regression.

Ann. Stat. 32, 407–499. doi: 10.1214/009053604000000067

Elsharkawy, M., Shimizu, M., Takahashi, H., and Hyakumachi, M. (2012). The plant growth-promoting fungus Fusarium equiseti and the arbuscular mycorrhizal fungus Glomus mosseae induce systemic resistance against Cucumber mosaic virus in cucumber plants. *Plant Soil* 361, 397–409. doi: 10.1007/s11104-012-1255-y

Farooq, A., Bukhari, S. A., Akram, N. A., Ashraf, M., Wijaya, L., Alyemeni, M. N., et al. (2020). Exogenously applied ascorbic acid-mediated changes in osmoprotection and oxidative defense system enhanced water stress tolerance in different cultivars of saower (carthamus tinctorious l.). *Plants* 9:104. doi: 10.3390/plants9010104

Gao, J. (2006). *The Experimental Guidance of Plant Physiology*. Beijing: Higher Education Press.

Gonzíalez-Moro, M., Gonzíalez-Moro, I., Estavillo, J. M., Aparicio-Tejo, P. M., Marino, D., et al. (2021). A multi-species analysis defines anaplerotic enzymes and amides as metabolic markers for ammonium nutrition. *Fronti. Plant Sci.* 11:2250. doi: 10.3389/fpls.2020.632285

Haddidi, I., Duc, N. H., Tonk, S., Rápó, E., and Posta, K. (2020). Defense enzymes in mycorrhizal tomato plants exposed to combined drought and heat stresses. *Agronomy* 10:1657. doi: 10.3390/agronomy10111657

Hesheng, L. (2000). *Principles and Techniques of Plant Physiological and Biochemical Experiments*. Beijing: Higher Education Press.

Islam, M., Matsui, T., and Yoshida, Y. (1996). Carbohydrate content and the activities of sucrose synthase, sucrose phosphate synthase and acid invertase in different tomato cultivars during fruit development. *Sci. Horticult.* 65, 125–136. doi: 10.1016/0304-4238(95)00871-3

Jaafar, H. Z. E., Ibrahim, M. H., and Mohamad Fakri, N. F. (2012). Impact of soil field water capacity on secondary metabolites, phenylalanine ammonia-lyase (PAL), Maliondialdehyde (MDA) and photosynthetic responses of malaysian kacip fatimah (labisia pumila benth). *Molecules* 17, 7305–7322. doi: 10.3390/molecules17067305

Kaur, S., Gupta, A. K., and Kaur, N. (2006). Effect of hydro- and osmopriming of chickpea (Cicer arietinum L.) seeds on enzymes of sucrose and nitrogen metabolism in nodules. *Plant Growth Regulat.* 49, 177–182. doi: 10.1007/s10725-006-9103-9

Lan, Y., Liu, M., Yan, Z., Xie, H., Shen, X., Zhang, L., and Lin, C. (2016). Comparison of fast breeding and regeneration seedling traits of salvia miltiorrhiza divinorum from different geographical sources. *Jiangsu Agric. Sci.* 44, 103–107. doi: 10.21273/HORTSCI13072-18

Lantz, B. (2013). *Machine Learning With R*. Birmingham: Packt Publishing Ltd.

Lea, P. J. and Morot-Gaudry, J.-F. (2001). *Plant Nitrogen*. Berlin; Heidelberg: Springer.

Li, H. S. (2000). *Principles and Techniques of Plant Physiological Biochemical Experiment*. Beijing: Higher Education Press.

Li, Y., Yang, Y., Hu, Y., Liu, H., He, M., Yang, Z., et al. (2019). DELLA and EDS1 form a feedback regulatory module to fine-tune plant growth-defense tradeoff in arabidopsis. *Mol. Plant* 12, 1485–1498. doi: 10.1016/j.molp.2019.07.006

Li, Z. (2018). *Salvia miltiorrhiza* Burge (Danshen): a golden herbal medicine in cardiovascular therapeutics. *Acta Pharmacol Sin.* 39, 802–824. doi: 10.1038/aps.2017.193

Liang, B., Ma, C., Zhang, Z., Wei, Z., Gao, T., Zhao, Q., et al. (2018). Long-term exogenous application of melatonin improves nutrient uptake fluxes in apple plants under moderate drought stress. *Environ. Exp. Bot.* 155, 650–661. doi: 10.1016/j.envexpbot.2018.08.016

Luo, H. (2018). Biomass accumulation, photosynthetic traits and root development of cotton as affected by Irrigation and nitrogen-fertilization. *Front. Plant Sci.* 9:14. doi: 10.3389/fpls.2018.00173

Ma, S., Zhang, D., Lou, H., Sun, L., and Ji, J. (2016). Evaluation of the anti-inflammatory activities of tanshinones isolated from salvia miltiorrhiza var. alba roots in thp-1 macrophages. *J. Ethnopharmacol.* 199, 193–199. doi: 10.1016/j.jep.2016.05.018

Moriguchi, T., Abe, K., Sanada, T., and Shohei, Y. (1992). Levels and role of sucrose synthase, sucrose-phosphate synthase, and acid invertase in sucrose accumulation in fruit of asian pear. *J. Am. Soc. Hortic. Sci.* 117, 274–278. doi: 10.21273/JASHS.117.2.274

Noor, E., Eden, E., Milo, R., and Alon, U. (2010). Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. *Mol. Cell* 39, 809–820. doi: 10.1016/j.molcel.2010.08.031

Polle, A., Otter, T., and Seifert, F. (1994). Apoplastic peroxidases and lignification in needles of norway spruce (Picea abies L.). *Plant Physiol.* 106, 53–60. doi: 10.1104/pp.106.1.53

Salehin, M., Li, B., Tang, M., Katz, E., Song, L., Ecker, J. R., et al. (2019). Auxin-sensitive aux/iaa proteins mediate drought tolerance in arabidopsis by regulating glucosinolate levels. *Nat. Commun.* 10, 1–9. doi: 10.1038/s41467-019-12002-1

Sarker, U., and Oba, S. (2020). The response of salinity stress-induced A. *tricolor* to growth, anatomy, physiology, non-enzymatic and enzymatic antioxidants. *Front. Plant Sci.* 11:559876. doi: 10.3389/fpls.2020.559876

Shi, M., Huang, F., Deng, C., Wang, Y., and Kai, G. (2019). Bioactivities, biosynthesis and biotechnological production of phenolic acids in salvia miltiorrhiza. *Crit. Rev. Food Sci. Nutr.* 59, 953–964. doi: 10.1080/10408398.2018.1474170

Strain, H. H., and Svec, W. A. (1966). "Extraction separation, estimation and isolation of the chlorophylls," in and GR (eds) *The Chlorophylls*, eds V. P. Vernon, and G. R. Svec (New York, NY: Academic Press). p. 21–66.

Su, C., Ming, Q.-L., Rahman, K., Han, T., and Qin, L.-P. (2015). *Salvia miltiorrhiza*: traditional medicinal uses, chemistry, and pharmacology. *Chin. J. Natural Med.* 13, 163–182. doi: 10.1016/S1875-5364(15)30002-9

Verma, A. K., Upadhyay, S. K., Verma, P. C., Solomon, S., and Singh, S. B. (2011). Functional analysis of sucrose phosphate synthase (sps) and sucrose synthase (ss) in sugarcane (saccharum) cultivars. *Plant Biol.* 13, 325–332. doi: 10.1111/j.1438-8677.2010.00379.x

Wang, L., Ma, R., Liu, C., Liu, H., Zhu, R., Guo, S., et al. (2017). *Salvia miltiorrhiza*: a potential red light to the development of cardiovascular diseases. *Curr. Pharm. Des.* 23, 1077–1097. doi: 10.2174/1381612822666161010105242

Wei, Y., Xiong, S., Zhang, Z., Meng, X., Wang, L., Zhang, X., et al. (2021). Localization, gene expression, and functions of glutamine synthetase isozymes in wheat grain (triticum aestivum l.). *Front. Plant Sci.* 12:37. doi: 10.3389/fpls.2021.580405

Ya-Li, L. (2018). *Effects of endophytic fungi on plant growth and metabolism of salvia miltiorrhiza bge* (MA thesis). Chengdu University of Traditional Chinese Medicine.

Yoneyama, Tadakatsu, Fujimori, Tamaki, Yanagisawa, Shuichi, Hase, Toshiharu, Suzuki, and Akira (2015). N-15 tracing studies on *in vitro* reactions of ferredoxin-dependent nitrite reductase and glutamate synthase using reconstituted electron donation systems. *Plant Cell Physiol.* 56, 1154–1161. doi: 10.1093/pcp/pcv039

Zaheer, I. E., Ali, S., Saleem, M. H., Imran, M., Alnusairi, G. S. H., Alharbi,B. M., et al. (2020). Role of iron-lysine on morpho-physiological traits and combating chromium toxicity in rapeseed (brassica napus l.) plants irrigated with different levels of tannery wastewater. *Plant Physiol. Biochem.* 155, 70–84. doi: 10.1016/j.plaphy.2020.07.034

Zhang, J., Khan, S., Sun, M., Gao, Z. Q., Liang, Y. F., Yang, Q. S., et al. (2018). Coordinated improvement of grain yield and protein content in dryland wheat by subsoiling and optimum planting density. *Appl. Ecol. Environ. Res.* 16, 7847–7866. doi: 10.15666/aeer/1606_78477866

Zhang, Z. (1990). *The Experimental Guidance of Plant Physiology*. Beijing: Higher Education Press.

Zhao, Y., Wei, Z., and Ma, F. (2003). Influence of ammoniacal nitrogen on sucrose synthase and sucrose phosphate synthase in sugar beet. *Sugar Crops China* 3, 1–5.

Zhou, B., Zhao, G., Zhu, Y., Chen, X., Zhang, N., Yang, J., et al. (2021). Protective effects of nicotinamide riboside on H_2O_2-induced oxidative damage in lens epithelial cells. *Curr. Eye Res.* 46, 961–970. doi: 10.1080/02713683.2020.1855662

Zhu, G., Deng, X., and Zuo, W. (1983). Determination of free proline in plants. *Plant Physiol. J.* 1, 37–39.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Plant Science

ORIGINAL RESEARCH
published: 04 January 2022
doi: 10.3389/fpls.2021.784780

Multiplexed Genome Editing *via* an RNA Polymerase II Promoter-Driven sgRNA Array in the Diatom *Phaeodactylum tricornutum*: Insights Into the Role of StLDP

*Yogesh Taparia†, Achintya Kumar Dolui†, Sammy Boussiba and Inna Khozin-Goldberg**

Microalgal Biotechnology Laboratory, French Associates Institute for Agriculture and Biotechnology of Drylands, The Jacob Blaustein Institutes for Desert Research, Ben-Gurion University of the Negev, Midreshet Ben-Gurion, Sede Boqer, Israel

OPEN ACCESS

Edited by:
Eric Marechal, UMR5168 Laboratoire de Physiologie Cellulaire Vegetale (LPCV), France

Reviewed by:
Helena Cruz de Carvalho, École Normale Supérieure, France
Fantao Kong, Dalian University of Technology, China

***Correspondence:**
Inna Khozin-Goldberg
khozin@bgu.ac.il

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Systems and Synthetic Biology, a section of the journal Frontiers in Plant Science

Received: *28 September 2021*
Accepted: *29 November 2021*
Published: *04 January 2022*

Citation:
Taparia Y, Dolui AK, Boussiba S and Khozin-Goldberg I (2022) Multiplexed Genome Editing via an RNA Polymerase II Promoter-Driven sgRNA Array in the Diatom Phaeodactylum tricornutum: Insights Into the Role of StLDP. Front. Plant Sci. 12:784780. doi: 10.3389/fpls.2021.784780

CRISPR/Cas9-mediated genome editing has been demonstrated in the model diatom *P. tricornutum*, yet the currently available genetic tools do not combine the various advantageous features into a single, easy-to-assemble, modular construct that would allow the multiplexed targeting and creation of marker-free genome-edited lines. In this report, we describe the construction of the first modular two-component transcriptional unit system expressing *Sp*Cas9 from a diatom episome, assembled using the Universal Loop plasmid kit for Golden Gate assembly. We compared the editing efficiency of two constructs with orthogonal promoter-terminator combinations targeting the StLDP gene, encoding the major lipid droplet protein of *P. tricornutum*. Multiplexed targeting of the StLDP gene was confirmed *via* PCR screening, and lines with homozygous deletions were isolated from primary exconjugants. An editing efficiency ranging from 6.7 to 13.8% was observed in the better performing construct. Selected gene-edited lines displayed growth impairment, altered morphology, and the formation of lipid droplets during nutrient-replete growth. Under nitrogen deprivation, oversized lipid droplets were observed; the recovery of cell proliferation and degradation of lipid droplets were impaired after nitrogen replenishment. The results are consistent with the key role played by StLDP in the regulation of lipid droplet size and lipid homeostasis.

Keywords: multiplexed genome editing, CRISPR, lipid droplet, Stramenopile-type lipid droplet protein, two-component transcriptional unit, RNA polymerase II promoter, *Phaeodactylum tricornutum*

INTRODUCTION

Diatoms are found in diverse environments, from polar to tropical, in both marine and freshwater bodies, in aquatic and soil ecosystems. Nearly 20,000 species are estimated to currently exist, with about 12,000 extant species documented in the scientific literature (Guiry, 2012). In ecosystems with high nutrient availability, diatoms rapidly multiply, followed by a collapse in

population, displaying typical "boom and bust" cycles. Due to their variety and vast numbers, diatoms are major contributors to the ocean's carbon, nitrogen, and silicon cycles (Benoiston et al., 2017). Within the marine environment, diatoms constitute a significant portion of phytoplankton, with estimated primary productivity of 45%, and they contribute nearly 20–40% of the global oxygen production. This success and dominance in several environments may be ascribed to their widespread horizontal gene transfer (Vancaester et al., 2020) and their distinct evolutionary origin from red algal secondary endosymbiosis (Armbrust, 2009).

To date, the study of diatom molecular physiology has been dominated by *Phaeodactylum tricornutum*, having a fully sequenced and annotated genome (Bowler et al., 2008) and a growing genetic manipulation toolkit (Apt et al., 1996; Zaslavskaia et al., 2000; Karas et al., 2015; Diner et al., 2016; Bielinski et al., 2017; Serif et al., 2017; Buck et al., 2018; Taparia et al., 2019; Pollak et al., 2020). Moreover, *P. tricornutum* has been used for developing several biotechnological products *via* either heterologous expression (Hempel et al., 2011) or the introduction of entire biosynthetic pathways using synthetic biology tools. Likewise, *P. tricornutum* has been engineered to produce high-value compounds, such as terpenoids (D'Adamo et al., 2018; Fabris et al., 2020) and vanillin (Slattery et al., 2018). *P. tricornutum* is a model oleaginous diatom whose lipid biosynthesis pathways have been intensively investigated. It has been commercially exploited as a source of high-value carotenoids (fucoxanthin) and omega-3 long-chain polyunsaturated fatty acids (LC-PUFA) and other lipids (Falciatore et al., 2020). A large number of genes implicated in the LC-PUFA biosynthesis and storage lipid metabolism have been characterized (Domergue et al., 2005; Hamilton et al., 2015; Haslam et al., 2020), to list a few.

Abiotic stress, such as nutrient deprivation (of nitrogen or phosphorus), triggers the accumulation of storage lipids (Sayanova et al., 2017) in the subcellular organelles, referred to as lipid droplets (LDs). The LD core of *P. tricornutum* is enriched in triacylglycerols (TAG) and enclosed by a monolayer of specific membrane lipids (Lupette et al., 2019). Considering the significance of LD-associated proteins in maintaining LD stability, lipid metabolism, and multifaceted interactions with other cellular compartments, proteomic analyses of isolated LDs have been reported in several studies, with the identification of the *bona fide* LD proteins in *P. tricornutum* and a few more oleaginous diatoms (Yoneda et al., 2016; Lupette et al., 2019; Leyland et al., 2020b). The structural integrity of LDs is deemed to be provided by proteins localized to or associated with the surface of LDs or embedded in the monolayer. However, the functional role of LD proteins in diatoms is yet to be established. One of the most abundant LD proteins, termed as a Stramenopile-type lipid droplet protein (StLDP; Phatr3_J48859), consistently appears in most of the proteomics studies of isolated LDs (Yoneda et al., 2016; Lupette et al., 2019; Leyland et al., 2020b). The amino acid sequence of StLDP features a central hydrophobic domain with conserved proline residues, which is assumed to assist in LD association. GFP-tagged StLDP localizes to LDs in N-deprived *P. tricornutum* (Shemesh, 2015; Yoneda et al., 2018). The overexpression of StLDP caused an increase in neutral lipid production under nitrogen (N) deprivation (Yoneda et al., 2018). Further study of StLDP function is warranted to address its possible role in LD metabolism and lipid homeostasis. However, successful metabolic engineering and functional characterization of candidate genes require efficient genome editing.

Initially, genome editing in *P. tricornutum* was successfully demonstrated *via* TALEN (Serif et al., 2017) and followed by CRISPR/Cas9 (Nymark et al., 2016). CRISPR/Cas9-mediated genome editing is a simple and versatile tool for creating targeted genome modifications and, as a result, has rapidly been adopted for several eukaryotic and prokaryotic species. CRISPR/Cas9 technology is especially valuable for asexual species, such as the model diatom *P. tricornutum*, for which some of the tools used in forward or reverse genetics cannot be applied. Moreover, RNAi for gene expression knock-down is known to be unstable in this diatom (Bielinski et al., 2017). The discovery of *E. coli*-mediated conjugation for the delivery of non-integrating extrachromosomal DNA (diatom episome) to the nucleus of *P. tricornutum* (Karas et al., 2015; Diner et al., 2016) presented an ideal system for delivering CRISPR/Cas9 for genome editing (Sharma et al., 2018; Slattery et al., 2018; Moosburner et al., 2020) without its integration into the nuclear genome. RNP-mediated multiplexed genome editing has also been reported in *P. tricornutum* (Serif et al., 2018). However, in addition to desired genome editing, this system requires the creation of auxotrophs as counterselection markers that require maintenance on media with special supplements. Moreover, most of these reports utilize special equipment for the delivery of gene editing constructs (biolistic gene gun) and expensive screening methods for identifying the resulting genome-editing events (RT-PCR for HRM assay or T7EI assay).

Recently, an episomal multiplexed CRISPR/Cas9 system was shown to use *Sp*Cas9 transcriptionally coupled with *Sh*Ble (selectable marker) to ensure the survival of only those exconjugants that expressed *Sp*Cas9 (Moosburner et al., 2020). Inspired by the successful report of *Sp*Cas9 transcriptionally coupled to the selectable marker, we further improved the system to allow modular assembly and the expression of an sgRNA array from an RNA polymerase II promoter (Mikami et al., 2017; Wang et al., 2018) to episomally express a multiplexed CRISPR/Cas9 system capable of creating genomic lesions that can be detected by a simple colony PCR. In this report, we demonstrate multiplexed targeting of the Stramenopile-type lipid droplet protein (StLDP; Yoneda et al., 2016) gene, leading to a large genomic lesion. Further, we were able to isolate the homozygous deletion lines of the StLDP gene for assessing the effect on phenotypes and traits related to LD biogenesis.

MATERIALS AND METHODS

Media Preparation and *P. tricornutum* Culture

P. tricornutum CCAP1055/1 (Pt 1) cultures were routinely diluted in a 1×RSE medium for maintenance in an exponential growth phase (Leyland et al., 2020b). *P. tricornutum* were cultured on

¼ × RSE + 1% agar +5% LB as the pre-conjugative and conjugation culture medium following the protocol described in (Diner et al. (2016).

Assembly of Level-1 Transcriptional Units and the Level-2 Diatom Episome

In the following Level-1-2 and Level-1-3 transcriptional units described, promoter-49202 + terminator-49202 and promoter-H4-1b + terminator FcpA combinations were constantly maintained, and these promoter-terminator combinations occurred orthogonally when assembled in a Level-2 diatom episome multiplexed CRISPR/Cas9 vector. Level-1 and Level-2 constructs were assembled using Golden Gate assembly reactions, as described in Pollak et al. (2019), using a pCA uLoop plasmid kit.

A Level-1_2 sgRNA array expression cassette was assembled using pL0-AC-Pr49202 or pL0-AC-PrH4-1b + pL0-HH-Ribozyme + sgRNA1 + sgRNA2 + sgRNA3 + sgRNA4 + pL0-HDV-Ribozyme + pL0-EF-Tr49202 or pL0-EF-TrFcpA (**Figure 1A**). Each sgRNA spacer fragment consisted of a complemetary forward and a reverse oligo, which were annealed and extended in a PCR reaction, of the form sgRNA-F aa**ggtct**caE1$(N)_{20}$GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATC and POS-R aa**ggtct**caE2aaAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTAT, where bold-face alphabets are the *BsaI* restriction site, E1 and E2 are unique Golden Gate overhangs (https://ggtools.neb.com/getset/run.cgi (Potapov et al., 2018; Pryor et al., 2020), (N)20 is a 20-bp genomic DNA targeting sequence selected from the CRISPOR web-tool (Haeussler et al., 2016; Concordet and Haeussler, 2018), and underlined sequences are complementary bases for annealing and extension in a PCR reaction. Assembly of a Level-1-2 sgRNA array was evaluated by PCR using HH-F (CCGTGAGGACGAAACGAGT) and HDV-R (CCGAAGCA TGTTGCCCAGC) primers. Details of the gRNA sequences targeting the StLDP coding sequence are listed in **Table 1**.

A Level-1_3 SpCas9 transcriptional unit was assembled using pL0-AC-Pr49202 or pL0-AC-PrH4-1b + pL0-CD-SpCas9 (PX459) + pL0-DB5-P2A + pL0-B5E-ShBle (or pL0-B5E-Bsr1) + pL0-EF-Tr49202 or pL0-EF-TrFcpA (**Figure 1B**). The SpCas9 sequence was amplified from Addgene plasmid #62988 (Ran et al., 2013). Level-1_1 consisted of a *StuI* + *PvuII* digested and religated episome to delete most of the *Sh*Ble cassette, and Level-1_4 consisted of a 24-bp spacer. Annotated nucleotide sequences for each part described above can be found in **Supplementary Data Sheet 2**.

The Level-2 diatom episome expression plasmid was constructed *via* a one-pot Golden Gate assembly reaction and transformed into NEB® Stable chemically competent *E. coli* and incubated for 20 h at 30°C. Individual *E. coli* colonies were patched on LB agar + Spectinomycin (30 mg/L) plates.

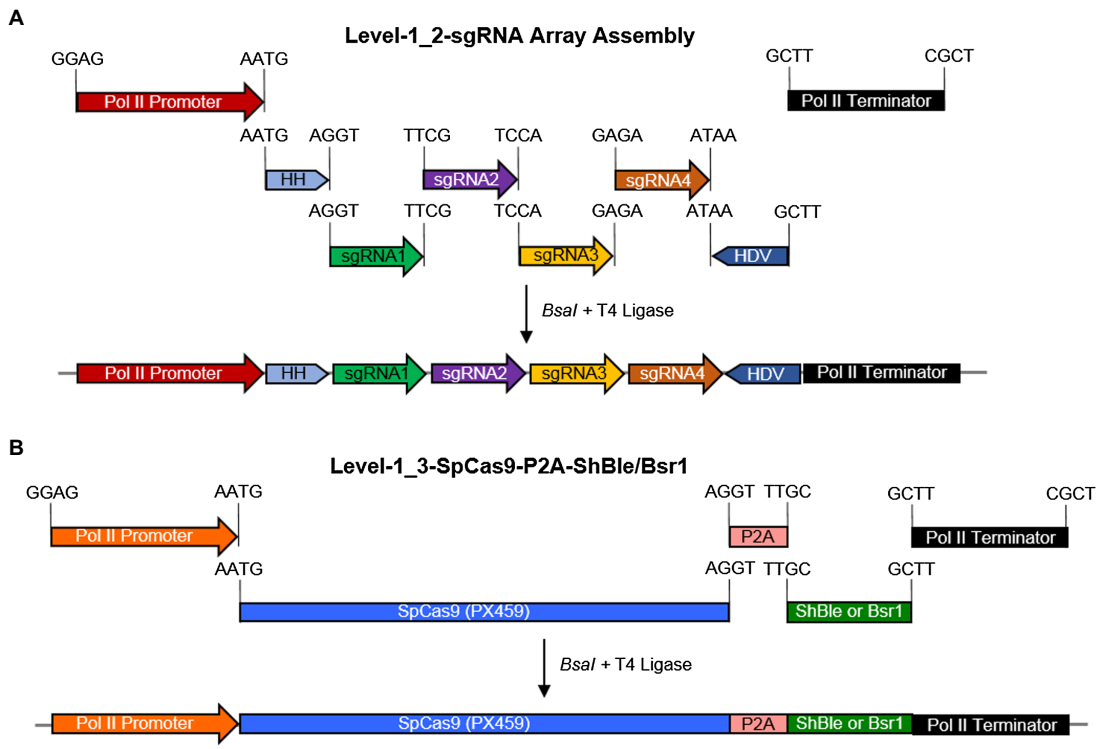


FIGURE 1 | A schematic diagram for the assembly of Level-0 Universal Loop parts into Level-1 transcriptional units and the final assembly of a Level-2 CRISPR/Cas9 diatom episome in the pCA plasmid kit. **(A)** a pCA-Level-1_2-sgRNA array was assembled in a Level-1 Golden Gate reaction with the *BsaI* enzyme and T4 DNA ligase for 30 cycles by combining Level-0 parts and sgRNA fragments with indicated overhangs. **(B)** A pCA-Level-1_3-SpCas9-P2A-*Sh*Ble/*Bsr1* transcriptional unit was assembled in a Level-1 Golden Gate reaction with the *BsaI* enzyme and T4 DNA ligase for 25 cycles by combining Level-0 parts with indicated overhangs. Promoter-terminator combination used for the sgRNA array is not reused for *Sp*Cas9-P2A-*Sh*Ble/*Bsr1* in the same Level-2 episome.

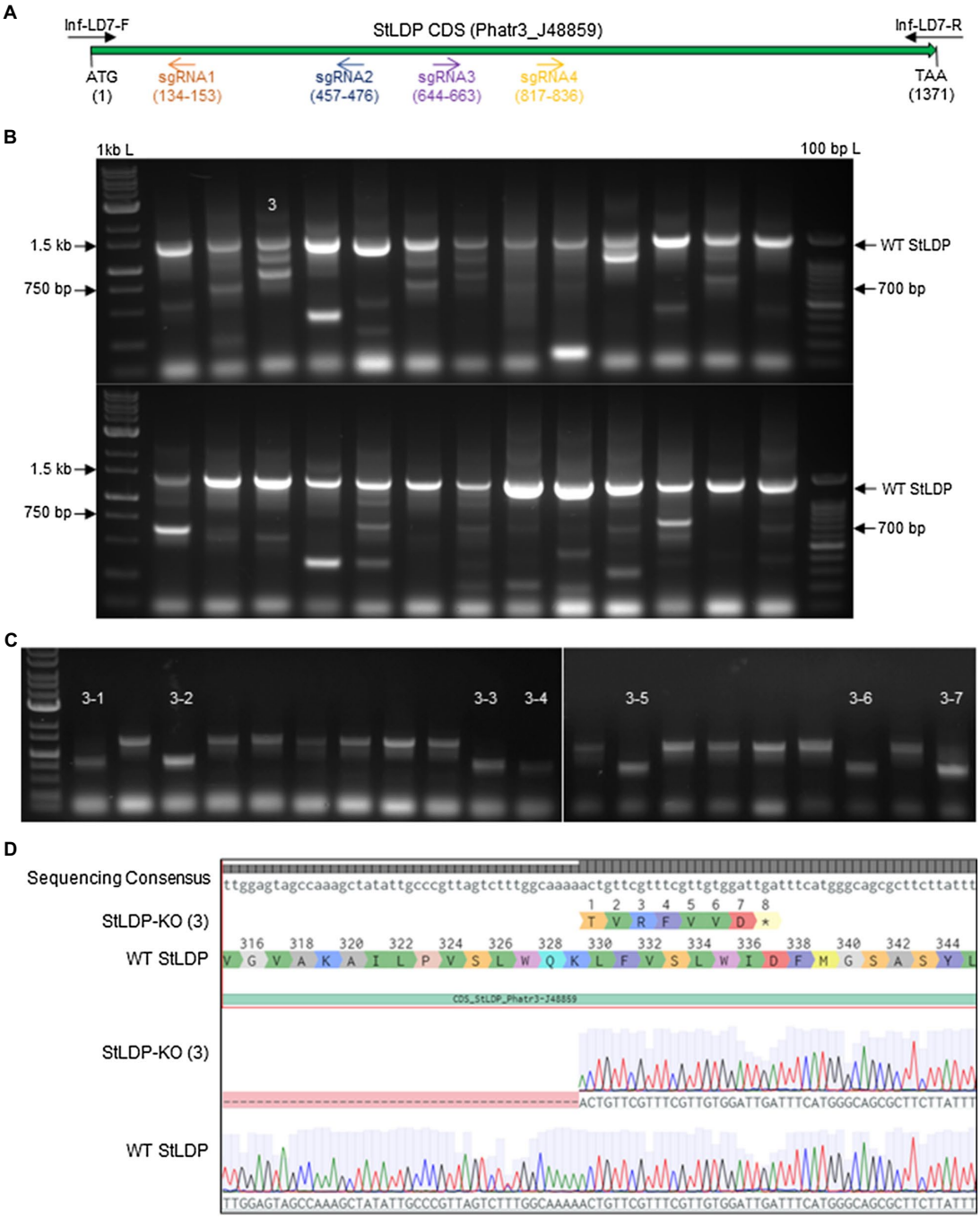


FIGURE 2 | PCR screening for gene editing in primary exconjugants in *P. tricornutum* Pt1 (CCAP1055/1). **(A)** A schematic of the StLDP gene (Phatr3_J48859), indicating positions of sgRNA-guided cleavage sites. **(B)** Twenty-six primary exconjugants with heterozygous or chimeric deletions in the StLDP gene are shown. Wt *StLDP*, amplified with INF-StLDP-F and INF-StLDP-R primers (Materials and Methods), produces a 1,400-bp product; amplicon below 700 bp indicates a deletion resulting from the multiplexed targeting of all four sgRNA. 1 kb L = GD 1 kb DNA Ladder RTU from GeneDireX Inc.; 100 bp L = 100 bp DNA Ladder RTU from GeneDireX Inc. **(C)** PCR screening of single colonies from primary exconjugant StLDP (3) to identify lines with homozygous deletion. **(D)** Partial sanger sequencing of the StLDP gene in primary exconjugant StLDP (3).

An on-colony PCR of patched *E. coli* colonies using HH-F 5′-CCGTGAGGACGAAACGAGT-3′ and HDV-R 5′-CCG AAGCATGTTGCCCAGC-3′ primers was performed to confirm the integrity of the sgRNA array prior to inoculating a culture for plasmid purification. All liquid cultures for *E. coli* (plasmid isolation and conjugation) were inoculated in LB24 medium (Wood et al., 2017) and grown at 30°C with appropriate antibiotics and shaken at 225 rpm.

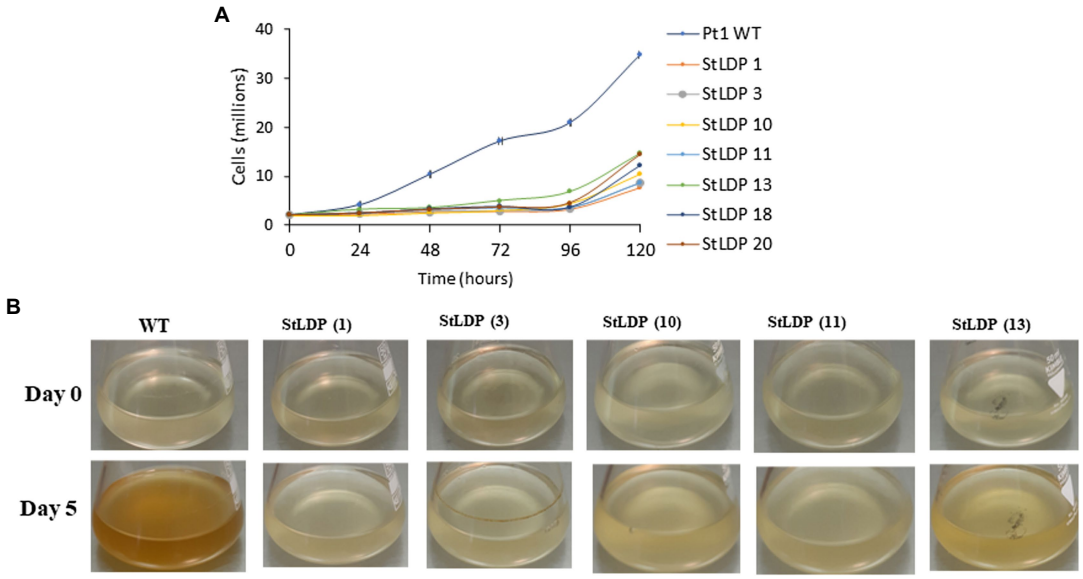


FIGURE 3 | Growth of wild-type Pt1 and *StLDP*-KO lines in the nutrient-replete medium. **(A)** Growth of the cultures monitored by cell counting. **(B)** Images of *P. tricornutum* cultures, representing relative growth from day 0 to 5 days of cultivation in the nutrient-replete medium. The initial cell density ~2×10^6 cells/ml. Wild-type and *StLDP*-KO lines were cultivated in an incubator shaker with a CO_2-enriched atmosphere (100 ml/min) at 22°C with a shaking speed of 120 rpm under a light intensity of 50 µmol photon m^{-2} s^{-1}.

Creation of a Conjugative *E. coli* Strain for Delivery of the Diatom Episome

NEB® Stable *E. coli*, transformed with pTA-Mob (Strand et al., 2014), served as a conjugative *E. coli* strain for delivery of a Level-2 expression episome to *P. tricornutum*. Chemically competent NEB® Stable *E. coli*, containing pTA-Mob, was transformed with a Level-2 diatom episome and selected on LB agar plates containing Gentamycin (20 mg/L) + Spectinomycin (30 mg/L) overnight at 30°C. Resultant *E. coli* colonies were patched and rescreened with PCR primers HH-F and HDV-R to ensure the stability of the diatom episome.

Conjugation and Selection

PCR-verified *E. coli* colonies were used for conjugation as described in Diner et al. (2016). Post-conjugation, *P. tricornutum* cells were selected on ¼ × RSE-agar plates containing Zeocin (100 mg/L) and Kanamycin (50 mg/L).

Detection of Genome-Edited Strains

Post-selection, exconjugant *P. tricornutum* colonies were screened via colony PCR using the primers INF-LD7-F (GCACATCGGGCTCGGGGTACCCTTCTTCGAGCAATCCGAA) + INF-LD7-R (TAATCATTTACTTAATTAGCTTGCAGGAACAAGCATG). Underlined sequences complementary to the StLDP coding sequence (Phatr3_J48859) were used to calculate the annealing temperature. An amplification product of 1403 bp indicated a lack of gene editing, and a 706 bp or lower DNA fragment indicated multiplexed genome editing.

Isolation of Homozygous Genome-Edited StLDP Lines

Putative chimeric/heterozygous exconjugants, with unedited and edited PCR bands of nearly equal intensity, were diluted and replated to obtain single colonies. These were patched, and a colony PCR was performed to identify homozygous genome-edited StLDP lines. One of the lines identified as a homozygous genome-edited *StLDP*-KO line was confirmed by Sanger sequencing (**Supplementary Figure 1**).

P. tricornutum Cultivation and Determination of Growth Parameters

Wild-type and homozygous genome-edited lines were cultivated in an incubator shaker at 22°C and a shaking speed of 120 rpm under a light intensity of 50 µmol photon m^{-2} s^{-1}, using a full-strength RSE medium. For the nitrogen deprivation experiments, cells from nutrient-replete cultures were harvested by centrifugation, washed in an N-free RSE medium, and resuspended in an N-free RSE medium at an initial cell density of 5×10^6 cell/ml. Cells were counted using an automated Luna cell counter (BioCat GmbH).

Fatty Acid Analysis

Fatty acid composition and content were determined as FA methyl esters (FAMEs), following the direct transmethylation of freeze-dried cell pellets and triacylglycerols, isolated from total lipid extracts. Samples placed in glass vials were transmethylated with 2% (v/v) sulfuric acid (H_2SO_4) in anhydrous methanol at 80°C for 1.5 h under an argon atmosphere with continuous stirring. Pentadecanoic acid (C_{15}:0; Sigma-Aldrich,

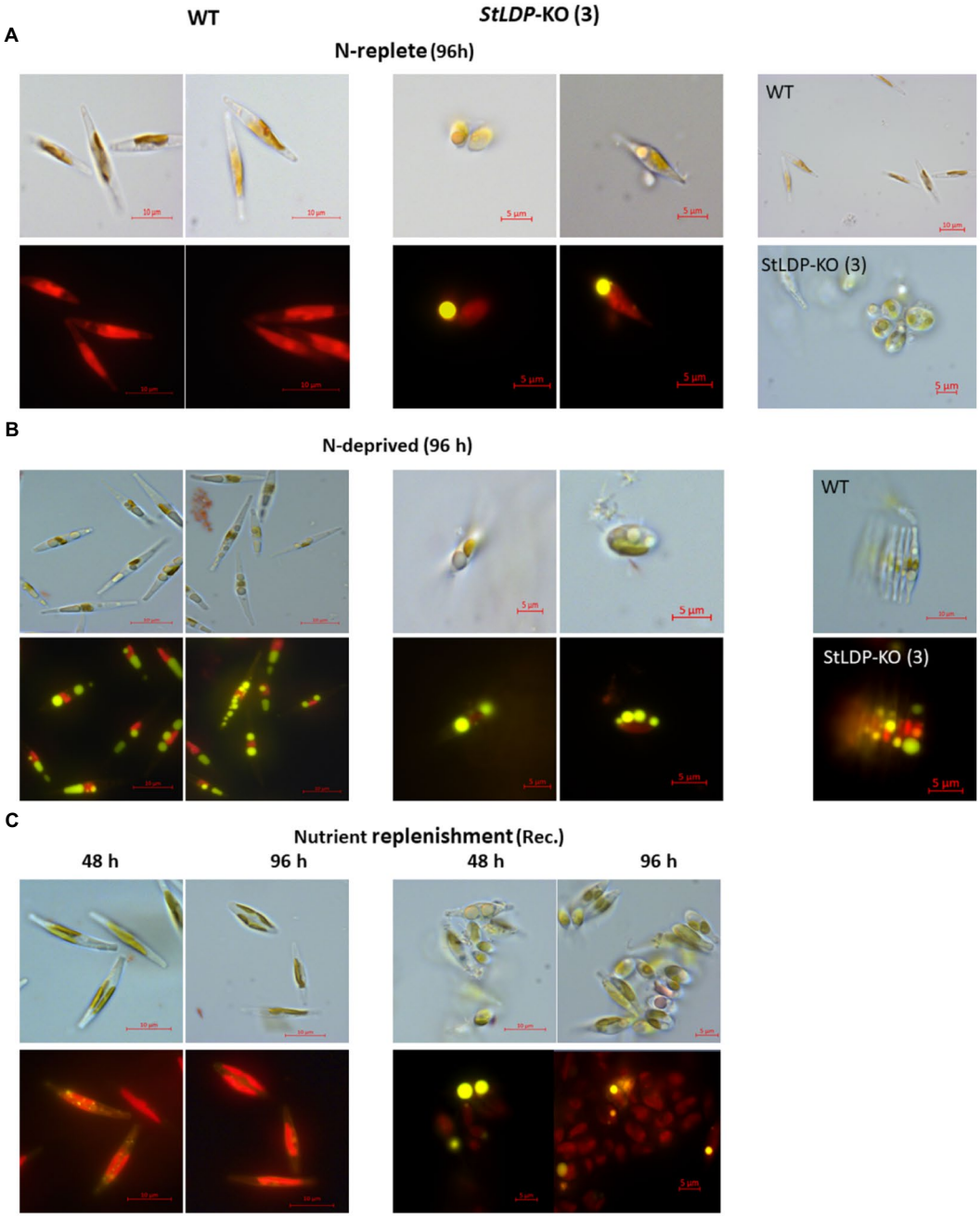


FIGURE 4 | Micrographs of Pt1 wild-type and *StLDP*-KO (3) cells under N-replete conditions **(A)**, N deprivation **(B)**, and following nutrient replenishment **(C)**. DIC micrographs and fluorescence micrographs of Nile Red stained cells (×1,000 magnification) are shown for each condition (indicated). Lipid droplets are visible as bright yellow fluorescent bodies. Pt1 wild-type and *StLDP*-KO (3) lines were cultivated in an incubator shaker with a CO_2-enriched atmosphere (100 ml/min) at 22°C with shaking speed of 120 rpm under a light intensity of 50 µmol photon m^{-2} s^{-1}.

United States; 0.5 mg/ml in stock solution) was added as an internal standard. The transmethylation reaction was terminated by cooling to room temperature and the addition of 1 ml of H_2O. FAMEs were extracted with hexane and quantified on a Trace GC Ultra (Thermo, Italy) equipped with an FID and a programmed temperature vaporizer (PTV) injector. The PTV injector was programmed to increase the temperature from 40°C at the time of injection to 300°C at sample transfer. The separation was achieved on a fused silica capillary column (SUPELCOWAX 10, Sigma-Aldrich, United States, 30 m × 0.32 mm) using the following oven temperature program: 1 min at 130°C, followed by a linear gradient to 220°C, and finally 10 min isocratic at 220°C. Helium was used as the carrier gas at a flow rate of 2.5 ml min^{-1}. The detector temperature

TABLE 1 | List of targets used for episome-mediated multiplexed editing of the StLDP gene.

Target	Cleavage position [a]-Strand_gRNA sequence	CFD score[b] (0–100)	Predicted efficiency[c] (0–100)	Off-target[d] sites
StLDP-1	131-R_GCGACTGGGTAAGTCGTACG	100	74	0
StLDP-2	454-R_TCTGGCACTACAGCGGACGA	100	69	0
StLDP-3	664-F_TTGCCATTTTCAATTCACAA	100	56	0
StLDP-4	837-F_TTGTACGCGAGCACGAACGC	100	70	0

[a]*Counted from ATG as first base.*
[b]*Cutting frequency determination (CFD) specificity score. A higher score correlates with a lower off-target effect in the genome (Doench et al., 2016).*
[c]*A higher score predicts a greater likelihood of cleavage for the given sequence (Doench et al., 2016).*
[d]*Number of off-target sites in the genome for a mismatch within the guide sequence (Concordet and Haeussler, 2018).*

was set at 280°C. FAMEs were identified by co-chromatography with commercial standards (Sigma-Aldrich, United States).

Lipid Extraction and Triacylglycerol Quantification

Freeze-dried cell pellets (150×10^6 cells) were treated with isopropanol, pre-heated to 80°C, for 10 min under an argon atmosphere. Isopropanol was removed and replaced with a mixture of methanol: chloroform (2:1, v/v). Samples were mixed for 1 h and centrifuged, and the solvent was collected and mixed with isopropanol. DDW was added to form two phases; the bottom phase was collected and evaporated under a stream of N_2 flow.

Triacylglycerols were separated by one-dimensional TLC on silica gel plates (Silica Gel 60, 10 × 10 cm, 0.25 mm thickness, Merck, Germany), using a solvent system of petroleum ether: diethyl ether: acetic acid (70:30:1, v/v/v). Lipid spots were visualized with iodine vapors, scraped from the plates, and transmethylated as described above.

Microscopy

Cells were observed under a Zeiss microscope (Carl Zeiss, Göttingen, Germany) equipped with an AxioCam digital camera, differential interference contrast (DIC) Nomarski optics, and epifluorescence. Lipid droplets were visualized using Nile Red staining: 1 µl of Nile Red (Sigma-Aldrich) solution in DMSO (100 µg/ml) was added to 10 µl of the sample, mixed by pipetting, and observed under a Filter Set 16 with a bandpass excitation at 485/20 nm and a long-pass emission beyond 515 nm.

RESULTS

Golden Gate Assembly Based a uLoop System Enables the Construction of a Highly Modular Genome-Editing Episome With Multiplexed sgRNA Capability

Golden Gate assembly combines a restriction-ligation cloning method into a one-pot reaction, which allows a seamless combination of interchangeable modules, thus achieving a combinatorial synthesis of expression vectors in parallel. Using an expanded syntax of the uLoop system, four sgRNA targets for the StLDP gene (Phatr3_J48859) were seamlessly assembled into a Level-1 transcriptional unit (**Figure 1A**). Similarly, a *Sp*Cas9 coding sequence was translationally coupled to the *Sh*Ble/*Bsr1* selectable marker *via* a P2A peptide with the facility to exchange promoter and terminator sequences as demanded by experiments or for future optimization (**Figure 1B**).

Stable Assembly and Delivery of a Level-2 Diatom Episome With Tandem sgRNA Repeats

While a multiplexed sgRNA transcriptional unit can be assembled into a Level-1_2 construct using routine cloning strains, such as NEB® Turbo, assembly with Level-1_3-*Sp*Cas9 into the final Level-2 expression episome fails without the use of a special *E. coli* strain, such as NEB® Stable, which is capable of maintaining DNA with multiple repeats. The use of LB24 broth (Wood et al., 2017) for culturing NEB® Stable *E. coli* at 30°C and 225 rpm ensured a comparable growth rate to EPI300™ *E. coli* grown at 37°C and 225 rpm, as used routinely for the conjugative delivery of the episome to *P. tricornutum*. To maintain episome stability, while culturing NEB® stable cells, LB agar plates and broth were incubated at 30°C. Slower growth in a liquid medium at 30°C was compensated by using a rationally balanced medium, such as LB24 [35].

Use of RNA Polymerase II Promoters Allows the Multiplexed Expression of an sgRNA Array

A tandem array of sgRNA targets, flanked by Hammer Head and Hepatitis Delta Virus ribozymes, was expressed as a single transcript using an RNA Pol-II promoter, precluding the use of multiple Pol-III promoter-terminator cassettes for each target and leading to a compact and efficient assembly of the CRISPR/Cas9 editing episome. Interspersing 6 bp filler sequences (5′-AA-3′ + 4 bp-GG overhang) between each sgRNA target allowed the endogenous ribonuclease processing of the tandem sgRNA array into active Cas9-RNP for genome editing.

Evaluation of Genome-Editing Events and Their Frequency With Two Constructs With Orthogonal RNA Polymerase II Promoters

Using four targets against the coding sequence of the StLDP gene of *P. tricornutum* (**Table 1**), we created a ~ 500 bp lesion detected by a simple on-colony PCR-based screening strategy (**Figure 2**). A direct on-colony PCR of *P. tricornutum* exconjugants allowed the rapid detection of genome-editing events for further study and evaluation (**Figure 2B**). The majority of edited events resulted from the construct in which *Sp*Cas9(PX459) and *Sh*Ble

TABLE 2 | Editing efficiency of two episome constructs with orthogonal RNA polymerase II promoters driving gRNA and Cas9 expression to detect deletions in the StLDP gene.

Promoter-terminator (sgRNA array/Cas9-P2A-ShBle)	Repetition	TotaleExconjugants	Edited events detected by PCR	Efficiency (%)
PrH4_1b-TrFcpA / Pr49202-Tr49202	1	80	9	11.3
	2	167	23	13.8
	3	120	8	6.7
Pr49202-Tr49202/PrH4_1b-TrFcpA	1	56	1	1.8
	2	38	1	2.6
	3	40	5	12.5

expression was driven by Pr49202-Tr49202, and the sgRNA array was transcribed by the PrH41b-TrFcpA promoter-terminator combination, which displayed editing at a frequency of 6.7–13.8% (**Table 2**). The sequences of the promoter and terminator combinations can be found in (Pollak et al., 2020). A PCR screening revealed that most primary exconjugants were either chimeric or heterozygous mutants (**Figure 2B**). The bands detected between the WT StLDP (~1,400 bp) and 700 bp (product of all four sgRNAs simultaneous deletion) are most likely deletions resulting from two or three sgRNAs.

Isolation of Homozygous Deletion StLDP Mutants

PCR-screened primary exconjugants, which displayed gene editing and the resultant PCR amplicons that were equal in intensity to the wild-type *StLDP* sequence, were diluted and replated. Lines with homozygous deletion were confirmed *via* PCR screening from the single colonies emerging from the replating (**Figure 2C**). Sanger sequencing from one of the lines, designated *StLDP*-KO (3), further confirmed the line to be a homozygous deletion event (**Supplementary Figure 1**; **Figure 2**). Zeocin selection pressure was maintained till the isolation of single colonies with a 500 bp homozygous deletion. In order to eliminate the background effect of Cas9 and conjugation, the lines were diluted several times under nonselective conditions to allow the CRISPR episome to be cured (Karas et al., 2015). These gene-edited lines (additionally confirmed by PCR on aliquots of liquid cultures; **Supplementary Figure 2**) free of the CRISPR/Cas9 episome were then compared to the WT under specified experimental conditions to evaluate the effect of disruption of the StLDP gene on LD formation.

Gene-Edited Lines Display Growth Phenotype and Accumulate Lipid Droplets in Replete Conditions

In this work, we report the initial characterization of the gene-edited lines in terms of growth, morphology, and fatty acid production under N-replete and N-deprived conditions. Seven transgenic *StLDP*-KO lines were monitored in the nutrient-replete medium for characterizing their growth performance (cell concentration) and microscopic analysis. All the examined gene-edited lines exhibited impaired growth and bleached phenotypes compared with the wild-type Pt1 (**Figures 3A,B**). All seven *StLDP*-KO lines exhibited virtually the same growth inhibition relative to the wild type (**Figure 3A**). Furthermore, light microscopy screening revealed that *StLDP*-KO lines formed LDs during the growth in the nutrient-replete medium. For a detailed microscopic analysis, we selected line *StLDP*-KO (3) with the 500 bp deletion confirmed by sequencing (**Supplementary Figure 1**). The morphology of *StLDP*-KO (3) cells grown in the replete medium differed markedly from the wild-type cells, dominated by fusiform cells. Three morphotypes of fusiform, circular, and oval cells were observed in the *StLDP*-KO (3) culture. In the wild-type cells grown in the N-replete medium, LDs, visualized by Nile Red staining, were either tiny or absent (**Figure 4A**), while *StLDP*-KO (3) cells produced large LDs (**Figure 4A**). Fusiform cells had one to two LDs with a diameter of 1.5–2 μm; however, multiple LDs were occasionally observed (**Figure 4A**). Oval and circular cells with a single LD were most abundant in the culture of *StLDP*-KO (3); these cells were prone to aggregation and the formation of stable clumps (**Figure 4A**). Because aggregation of oval and circular cells hindered cell counting and staining the LDs with Nile Red, intensive vortexing was required before the analysis.

Following the transfer to the N-deprived medium (**Figure 4B**), the cells of *StLDP*-KO (3) with pre-formed LDs did not display substantial morphological changes compared with the cells in nutrient-replete medium; furthermore, cell clumping increased. Circular LDs occupied most of the cell volume, while the plastid and other compartments seemed to be substantially abridged. Cell clumping decreased after nutrient replenishment (**Figure 4C**); however, clumps were observed after 96 h of recovery. Within 48 h following nutrient resupply, LDs disappeared from wild-type cells, whereas large, circular LDs were observed in the *StLDP*-KO (3) cells.

Gene-Edited Lines Accumulate TAG in Replete Conditions

To examine whether and how the lesion in the *StLDP* gene affected the production and the relative abundance of fatty acids (% of total) under N-replete and N-deprived conditions and following nutrient replenishment, a FAMEs analysis was performed. Under nutrient-replete conditions, the fatty acid profile of the *StLDP*-KO (3) line showed a reduction in the proportion of the major omega-3 LC-PUFA, eicosapentaenoic acid, 20:5*n*3, as well as a noticeable increase in the proportion of palmitoleic acid (16:1*n*7) compared with the wild type (**Figure 5A**). The cells of the *StLDP*-KO (3) accumulated *ca.*

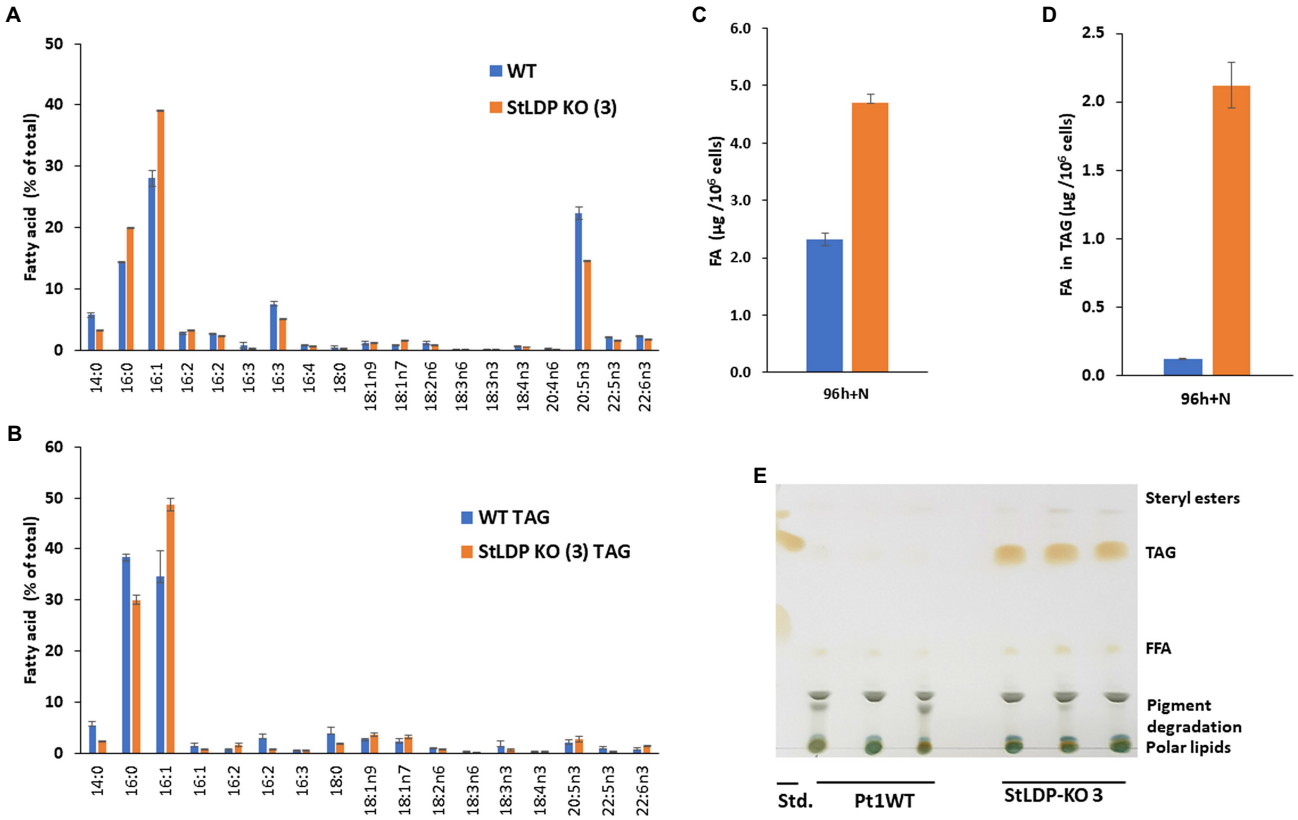


FIGURE 5 | Fatty acid analysis of wild-type Pt1 and *StLDP*-KO (3) under nutrient-replete conditions by gas chromatography (GC). **(A)** Relative abundance of total fatty acids (%) after 96 h of cultivation in a nutrient-replete medium. **(B)** Total fatty acid content per 10^6 cells. **(C)** Relative abundance of fatty acids in TAG (%). **(D)** Fatty acid content in TAG per 10^6 cells. **(E)** Separation of neutral lipids by TLC in a solvent system used for TAG isolation.

2.5-fold more fatty acids compared with the wild type under nutrient replete conditions (**Figure 5C**).

Since the *StLDP*-KO (3) line exhibited LD formation in replete conditions, we isolated TAG from the total lipid extract by one-dimensional TLC and subsequently quantified acyl groups accumulated in TAG as FAMEs. As was evident from the imaging of the TLC plate after exposure to iodine vapors, the *StLDP*-KO (3) accumulated TAG under replete conditions, while TAG formation was minute in the wild-type cells (**Figures 5D,E**). Quantification of fatty acids confirmed a substantial accumulation of TAG in the *StLDP*-KO (3) cells under nutrient-replete conditions. The main differences in the fatty acid profile of TAG comprised the increased percentage of 16:1 and a decrease of 14:0 and 16:0 compared with the TAG of the wild type (**Figure 5B**).

Under N deprivation, the main difference in the fatty acid profile comprised a decrease in the proportion of 16:0 with a concomitant increase in 16:1*n*7 (**Figure 6A**). In the course of nitrogen deprivation, the TFA content markedly increased in the wild type to 4.8 µg/10^6 cells (**Figure 6B**). In contrast, there was a relatively mild increase in the *StLDP*-KO (3) line (to *ca.* 6.0 µg/10^6 cells), compared with the N-replete cells (**Figure 6B**). TLC analysis performed to quantify TAG produced under N deprivation indicated that TAG increased less in the KO line than in the wild type (**Figure 6C**). Yet, the TAG content was higher in the *StLDP*-KO (3) cells. Following the resupply of nutrients to N-deprived cells, in line with the growth (**Figure 6A**) and microscopy data (**Figure 4C**), the growth recovery and the decrease in fatty acid content (per 10^6 cells) was impeded in *StLDP*-KO (3). In the wild type, the content of total fatty acids drastically decreased within 48 h, from 5 to 1.5 µg/10^6 cells, followed by an increase to ~3.0 µg/10^6 cells over the next 48 h (**Figure 6B**). In contrast, the total fatty acid content remained substantially higher in the *StLDP*-KO (3), but there was a steady, though relatively mild, decrease in the *StLDP*-KO (3). The fatty acid profile comprised ~40% of 16:1 and 20% of 16:0 in the *StLDP*-KO after 48 h of nutrient resupply (**Figure 6D**). In the wild type, these fatty acids comprised ~20 and 15%, respectively after 48 h of nutrient resupply. Since 16:0 and 16:1 are the major constituents of TAG in *P. tricornutum* (**Figure 5E**), this result may corroborate a delayed hydrolysis of TAG and LD breakdown in *StLDP*-KO (3) cells. Another indication of the impaired recovery of *StLDP*-KO (3) cells from N deprivation was the reduced proportion of the EPA of total fatty acids. By 96 h, the effect was moderate, as indicated by an increase in the percentage of two PUFAs, 16:3 and 20:5*n*3, associated with plastid membrane lipids.

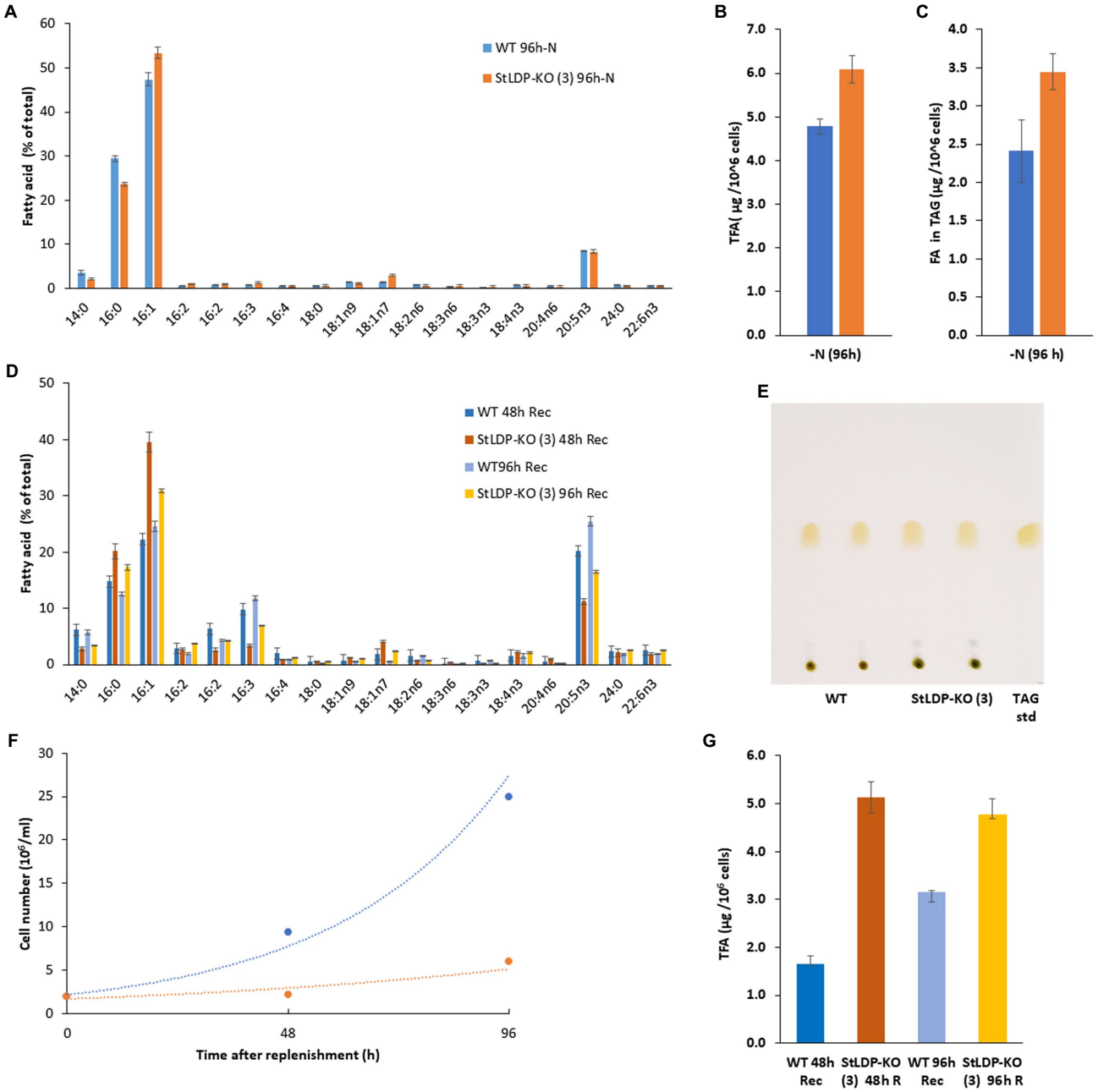


FIGURE 6 | Comparative analysis of wild-type Pt1 and *StLDP*-KO (3) under N deprivation and following nutrient replenishment. **(A)** Relative abundance of fatty acids (% of total) after 96 h of cultivation in the N-deprived medium. **(B)** Total fatty acid content (µg/10^6 cells) after 96h of N deprivation. **(C)** Fatty acid content in TAG (µg/10^6 cells) after 96h of N deprivation. The initial cell density $\sim 5 \times 10^6$ cells/ml. **(D)** Relative abundance of fatty acids (% of total) during recovery from N-deprivation. **(E)** Isolation of TAG from N -deprived cells by TLC: total lipids were loaded based on the equal cell number. **(F)** Cell concentration dynamics after nutrient replenishment. The initial cell density $\sim 2 \times 10^6$ cells/ml. **(G)** Relative abundance of fatty acids (% of total) 48 and 96 h after nutrient replenishment.

DISCUSSION

Although several incremental improvements have been achieved in deploying CRISPR/Cas9 for genome editing in *P. tricornutum* (Nymark et al., 2016; Kroth et al., 2018; Serif et al., 2018; Sharma et al., 2018; Stukenberg et al., 2018; Moosburner et al., 2020), none of these reports, to date, have incorporated the latest developments in multiplexed genome targeting as a modular system allowing rapid construct assembly for creating large genomic lesions that can be easily detected by PCR. Here, we described a scheme for modular assembly of a multiplexed genome-editing episome, based on Universal Loop assembly (Pollak, 2019; Pollak et al., 2019, 2020), to enable a simple PCR-based screening of edited lines, free of selection markers.

Multiplexed genome targeting allows the simultaneous editing of two or more loci or the creation of a large deletion that can be detected easily by routine PCR screening. Several strategies have been described in other model systems for multiplexing sgRNA expression from constructs that include stacked RNA polymerase III sgRNA expression cassettes (Moosburner et al., 2020), tRNA-sgRNA arrays expressed from RNA polymerase III promoters (Xie et al., 2015), and ribozyme-gRNA-ribozyme (RGR) arrays transcribed from RNA polymerase II/III promoters (Gao and Zhao, 2014).

The majority of multiplexed sgRNA expression systems utilize snRNA/ U6/U3 (RNA polymerase III) promoters, which are inherently inefficient at transcribing a long array of sgRNAs, typically interspersed with tRNA or ribozymes or Csy4-28 bp linkers (Hsieh-Feng and Yang, 2019). This limitation can be overcome by using RNA polymerase II promoters, which can easily transcribe long sgRNA arrays containing ribozymes for efficient cleavage of the array-transcript into mature sgRNAs. An efficient TCTU system for multiplexed gene editing in rice utilizing an RNA polymerase II promoter for transcribing an RGR array, was recently described, allowing simultaneous editing in three targets (Wang et al., 2018). Furthermore, in rice (Mikami et al., 2017), demonstrated that ribozymes are optional for processing the SpCas9-sgRNA transcript into functional RNP due to the presence of endogenous ribonucleases.

In this study, we combined the use of an RNA polymerase II-based TCTU approach to CRISPR/Cas9 from (Wang et al., 2018), the use of single HH and HDV ribozymes for the entire four target sgRNA array from (Gao and Zhao, 2014), an endogenous sgRNA processing capability from (Mikami et al., 2017), and a selectable marker transcriptionally coupled to SpCas9(PX459) expression (Moosburner et al., 2020) into an efficient, compact, episomally expressed CRISPR/Cas9 genome-editing system with multiplexed capability (**Figure 1**). Although the editing efficiency reported here is lower than that reported in other studies, our estimates are more conservative due to the fact that they are based on an overall observation of a large genic deletion rather than on an sgRNA-by-sgRNA basis. Moreover, the editing detection strategy used in this report is simple, requiring no special equipment or additional reagents in contrast to the methods, such as HRM or T7EI, used in other reports.

Recently, a Universal Loop kit (Pollak et al., 2019, 2020), a Golden Gate assembly-based cloning system, was released for several photosynthetic microalgae, including *P. tricornutum*. Such a Golden Gate cloning-based system offers an efficient, one-pot assembly of DNA parts with standardized overhangs, up to 24 parts (Potapov et al., 2018; Pryor et al., 2020). Two strong promoter-terminator combinations, PrH4_1b-TrFcpA (Slattery et al., 2018) and Pr49202-Tr49202 (Bielinski et al., 2017), were selected from the kit to construct a TCTU system. These were orthogonally combined either to transcribe the ribozyme-flanked sgRNA array or to express the SpCas9-P2A-ShBle cassette into two diatom episome constructs for testing editing efficiency. Surprisingly, of the two constructs, in the combination where the promoter H4_1b-terminator FcpA was used for expressing the sgRNA array and the promoter 49,202-terminator 49,202 was used for expressing SpCas9-P2A-*Sh*Ble, the number of exconjugants and gene editing was observed at a greater frequency. Moreover, most of the deletions observed were either chimeric or heterozygous. This highlights the need for testing more promoters, preferably from genes involved in cell division or circadian rhythm. Indeed, a recent study in *Chlamydomonas reinhardtii* reported an increase in homology-directed repair events or non-homologous end-joining events based on the time point within the cell cycle, when circadian-rhythm-synchronized cultures were delivered to Cas9 RNP (Angstenberger et al., 2020). Our study further suggests the feasibility of deploying a single transcriptional unit CRISPR/Cas9 without the use of ribozymes for multiplexed genome targeting. Such improvements in multiplexed genome-targeting systems hold promise for efficient transcriptional modulation or repression *via* CRISPRa and CRISPRi, especially for a model diatom species, such as *P. tricornutum*, allowing a thorough investigation of metabolic pathways and further opening avenues for their engineering.

Delivery of CRISPR/Cas9 has been achieved in *P. tricornutum* *via* the biolistic transformation of plasmids (Nymark et al., 2016; Sharma et al., 2018) or RNP (Serif et al., 2018) and *E. coli*-mediated conjugation (Sharma et al., 2018; Slattery et al., 2018; Stukenberg et al., 2018; Moosburner et al., 2020). Of all these methods, *E. coli*-mediated conjugation requires no specialized equipment or chemicals to maintain auxotrophs (for RNP counterselection) and can be performed in a high-throughput format. Additionally, the diatom episome can be cured upon the removal of selection pressure. While the assembly of constructs with one or two sgRNA targets is easily achieved, the addition of more targets for multiplexing increases repetitive DNA in plasmids and makes the construct unstable in *E. coli*. In our experiments, a Level-2 multiplex sgRNA episome could only be assembled in NEB® Stable *E. coli* specialized for maintaining plasmids with repetitive DNA. However, the assembly of a Level-1-2 sgRNA transcriptional unit with four targets could easily be achieved with regular *E. coli* strains, such as NEB® Turbo. Similarly, for the *E. coli*-mediated delivery of the multiplexed CRISPR/Cas9 episome, NEB® Stable *E. coli* was transformed with pTA-Mob (Strand et al., 2014) and the Level-2 diatom episome.

Multiplexed genome editing of the StLDP gene successfully created targeted and stable gene knockouts in *P. tricornutum* cells, carrying the genic lesion in the gene encoding StLDP and showing a strong growth impairment. *StLDP*-KO cells adopted altered shapes and morphologies distinct from the typical fusiform cells of the Pt1 wild type, suggesting a possible role of StLDP or the downstream effect of its KO in the cell morphology regulation. Furthermore, oval and circular cells were amply present in cultures of the edited lines, and they formed cell clumps. It is known that *P. tricornutum* cells secrete exopolymeric substances (EPS) with adhesive properties, particularly the oval cells (Willis et al., 2013; Galas et al., 2021). It has also been suggested that oval cells represent a form that is more resilient to stresses, in Galas et al. (2021) and references therein.

Furthermore, the StLDP gene KO resulted in LD formation and TAG accumulation in nitrogen-replete conditions. While the molecular mechanisms of LD formation in the N-replete cells of the gene-edited lines warrant further investigation, we can provide several plausible explanations. We speculate that the ablation of StLDP and the resulting changes in the assemblage of LD-associated proteins may disturb LD turnover in the replete cells followed by the coalescence of forming LDs. Generally diminutive or absent from the replete cells, LDs and the storage lipids sequestered within LDs undergo continuous turnover, as corroborated by knocking down the genes for major TAG lipases in two diatoms (Trentacoste et al., 2013; Barka et al., 2016). The study of the peculiar dynamics of LD formation by *P. tricornutum* strain Pt1 (Jaussaud et al., 2020) revealed a complex mode of LD size control and the lack of LD fusion in the presence of StLDP under N deprivation.

Furthermore, since the nature and composition of major lipid droplet proteins vary among different species, the downstream effect of knockdown/knockout of genes encoding proteins on lipid metabolism might be species- and taxon-dependent (Yoneda et al., 2016). For example, in the green microalga *C. reinhardtii*, RNAi-mediated silencing of *MLDP* led to an increase in the diameter of LDs in N-deprived conditions, without, however, affecting the TAG content (Moellering and Benning, 2010). Cultivation of *StLDP*-KO cells in the N-deprived medium caused detrimental effects on cell morphology (agglomerated circular and oval cells with oversized LDs) and did not cause a substantial increment in the TAG content per cell. Following nutrient replenishment, the recovery of *StLDP*-KO cells was severely impaired. The remobilization of storage lipids in the *StLDP*-KO during recovery was delayed, although not completely impaired, and coincided with at least partial recovery of the membrane systems, as indicated by microscopy and FA profiling (an increase in 16:3 and 20:5). The delay in TAG degradation could be attributed to the fact that the reduced surface area per volume of LDs in StLDP led to decreased TAG hydrolysis by lipases, acting in LD turnover at the LD surface. For example, in *Arabidopsis thaliana*, decreased lipid remobilization was observed when a gene for oleosin 1, a lipid droplet resident protein, was knocked out (Siloto et al., 2006). Consistent with our observation of the formation of oversized LDs in the *StLDP*-KO line, enlarged LDs, which did not cease to coalesce, were observed in the *mldp* lines of *C. reinhardtii* under prolonged N deprivation (Tsai et al., 2015). However, in the *StLDP-KO* line LDs were formed and TAG accumulated in replete cells, indicating a significant impact on TAG turnover. Further, the reduction of MLDP in *C. reinhardtii* caused a delay in TAG breakdown, which is consistent with our results. Therefore, our study indicates the important role played by StLDP in regulating LD size and number in *P. tricornutum* and suggests a possible function in stabilizing the LD surface. A more detailed study of the *StLDP*-KO lines is further required to reveal the role of this major LD protein in *P. tricornutum*.

Lipid droplets can also be degraded by the autophagy-related lipophagy process (Zienkiewicz and Zienkiewicz, 2020). It has recently been shown that the major LD surface protein of the eustigmatophyte *Nannochloropsis oceanica* interacts with the hallmark autophagy protein ATG8 during the autophagic catabolism of LDs (Zienkiewicz et al., 2020). The use of fluorescent-protein-tagged ATG8 helped to reveal contacts between LDs and autophagic structures in *C. reinhardtii* (Tran et al., 2019). A similar mechanism may operate in *P. tricornutum*, as suggested by the presence of the ATG8-interacting motif in StLDP (Leyland et al., 2020a). Therefore, it is possible to speculate that the ablation of StLDP disturbs autophagic degradation of LDs in nutrient-replete conditions, consistent with the augmented formation of LDs in replete cells and the impeded degradation following nutrient replenishment. These suggestions and other possible mechanisms and effects of StLDP ablation on cellular and LD proteomes and lipidomes will be experimentally tested in future research.

CONCLUSION

In this work, we have demonstrated the creation of a large homozygous deletion in the StLDP gene *via* a multiplexed CRISPR/Cas9 array expressed from the diatom episome. This report describes the essential cloning steps and *E. coli* strains required for successful accomplishment of multiplexed genome editing. We demonstrate here the feasibility of efficient multiplexed gene editing in a TCTU system *via* the diatom episome. Multiplexed targeting *via* RNA polymerase II promoters greatly reduced the size of the episome construct and overcame the necessity to find new snRNA/RNA polymerase III promoters for target stacking in a construct, especially for a recent model system. *E. coli*-mediated episome delivery of CRISPR/Cas9 to diatoms can be cured of exconjugants creating non-GM edited lines, which are desirable in commercial cultivation. Moreover, the US has recently deregulated gene editing in crops for creating single base changes or several base-pair deletions that can be achieved by traditional breeding practices (Stokstad, 2020). Our work also suggests that sgRNA arrays can be efficiently processed into functional sgRNA by endogenous ribonucleases in *P. tricornutum*, allowing the accommodation of more sgRNA targets instead of ribonucleases. Deploying four targets may provide a greater chance of ensuring generation of mutants with large deletion within a given genomic region. Moreover, four targets may be further used for creating double mutants in a single experiment. We hope that the Level-0 parts generated in this study (**Supplementary Material**) will be useful within the *P. tricornutum* community for further optimization of CRISPR/Cas9-mediated gene editing. The effectiveness of the method was exemplified by creating mutants of the major lipid droplet protein StLDP, allowing novel insights into its role in LD metabolism.

DATA AVAILABILITY STATEMENT

All sequences and data relevant to reproducing this work has been included in the article and its **Supplementary Material**.

AUTHOR CONTRIBUTIONS

YT and IK-G: conceptualization. YT and AKD: methodology. YT, AKD, and IK-G: formal analysis and writing – original draft preparation. IK-G: writing – review and editing and funding acquisition. IK-G and SB: supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Israel Science Foundation (grant no. 1958/18).

ACKNOWLEDGMENTS

YT and AKD acknowledge the scholarships from the Kreitman School of Advanced Graduate Studies (Ben-Gurion University). We thank Iftach Nachman for sharing the SpCas9-PX459 plasmid; Igor Ulitsky (Weizmann Institute of Science, Rehovot) for sharing the NEB® Stable *E. coli*; Tsafi Danieli (The Hebrew University of Jerusalem) for recommending the use of NEB® Stable *E. coli* for assembling unstable plasmids with multiple repeats; Vandana Bharadia and Tom Biton (Weizmann Institute of Science, Rehovot) for their help in procuring some of the materials for this work; Boris Zorin for helpful comments; and Samara Bel for the English language editing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2021.784780/full#supplementary-material

Supplementary Figure 1 | The entire Sanger sequencing of the edited StLDP gene in primary exconjugant StLDP-KO (3). The minimized figure is presented in **Figure 2D**.

Supplementary Figure 2 | Wt StLDP gene, amplified with INF-StLDP-F and INF-StLDP-R primers (Materials and Methods), produces a 1,400-bp product; the amplicon of ~700 bp in StLDP-KO (3) indicates a deletion resulting from the multiplexed targeting of all four sgRNA. The positions of DNA size markers are indicated.

REFERENCES

Angstenberger, M., De Signori, F., Vecchi, V., Dall'Osto, L., and Bassi, R. (2020). Cell synchronization enhances nuclear transformation and genome editing via Cas9 enabling homologous recombination in *Chlamydomonas reinhardtii*. *ACS Synth. Biol.* 9, 2840–2850. doi: 10.1021/acssynbio.0c00390

Apt, K. E., Grossman, A. R., and Kroth-Pancic, P. G. (1996). Stable nuclear transformation of the diatom *Phaeodactylum tricornutum*. *Mol. Gen. Genet.* 252, 572–579. doi: 10.1007/BF02172403

Armbrust, E. V. (2009). The life of diatoms in the world's oceans. *Nature* 459, 185–192. doi: 10.1038/nature08057

Barka, F., Angstenberger, M., Ahrendt, T., Lorenzen, W., Bode, H. B., and Büchel, C. (2016). Identification of a triacylglycerol lipase in the diatom *Phaeodactylum tricornutum*. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1861, 239–248. doi: 10.1016/j.bbalip.2015.12.023

Benoiston, A.-S., Ibarbalz, F. M., Bittner, L., Guidi, L., Jahn, O., Dutkiewicz, S., et al. (2017). The evolution of diatoms and their biogeochemical functions. *Philos. Trans. R. Soc. B: Biol. Sci.* 372:20160397. doi: 10.1098/rstb.2016.0397

Bielinski, V. A., Bolt, T. M., Dupont, C. L., and Weyman, P. D. (2017). Episomal tools for RNAi in the diatom *Phaeodactylum tricornutum*. *PeerJ* 5:e2907v1. doi: 10.7287/peerj.preprints.2907v1

Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., et al. (2008). The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* 456, 239–244. doi: 10.1038/nature07410

Buck, J. M., Río Bártulos, C., Gruber, A., and Kroth, P. G. (2018). Blasticidin-S deaminase, a new selection marker for genetic transformation of the diatom *Phaeodactylum tricornutum*. *PeerJ* 6:e5884. doi: 10.7717/peerj.5884

Concordet, J. P., and Haeussler, M. (2018). CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res.* 46, W242–W245. doi: 10.1093/nar/gky354

D'Adamo, S., Schiano di Visconte, G., Lowe, G., Szaub-Newton, J., Beacham, T., Landels, A., et al. (2018). Engineering the unicellular alga *Phaeodactylum tricornutum* for high-value plant triterpenoid production. *Plant Biotechnol. J.* 17, 75–87. doi: 10.1111/pbi.12948

Diner, R. E., Bielinski, V. A., Dupont, C. L., Allen, A. E., and Weyman, P. D. (2016). Refinement of the diatom episome maintenance sequence and improvement of conjugation-based DNA delivery methods. *Front. Bioeng. Biotechnol.* 4:65. doi: 10.3389/fbioe.2016.00065

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191. doi: 10.1038/nbt.3437

Domergue, F., Abbadi, A., Zähringer, U., Moreau, H., and Heinz, E. (2005). In vivo characterization of the first acyl-CoA Delta6-desaturase from a member of the plant kingdom, the microalga *Ostreococcus tauri*. *Biochem. J.* 389, 483–490. doi: 10.1042/BJ20050111

Fabris, M., George, J., Kuzhiumparambil, U., Lawson, C. A., Jaramillo-Madrid, A. C., Abbriano, R. M., et al. (2020). Extrachromosomal genetic engineering of the marine diatom *Phaeodactylum tricornutum* enables the heterologous production of monoterpenoids. *ACS Synth. Biol.* 9, 598–612. doi: 10.1021/acssynbio.9b00455

Falciatore, A., Jaubert, M., Bouly, J. P., Bailleul, B., and Mock, T. (2020). Diatom molecular research comes of age: model species for studying phytoplankton biology and diversity. *Plant Cell* 32, 547–572. doi: 10.1105/tpc.19.00158

Galas, L., Burel, C., Schapman, D., Ropitaux, M., Bernard, S., Bénard, M., et al. (2021). Comparative structural and functional analyses of the fusiform, oval, and triradiate morphotypes of *Phaeodactylum tricornutum* Pt3 strain. *Front. Plant Sci.* 12:584. doi: 10.3389/fpls.2021.638181

Gao, Y., and Zhao, Y. (2014). Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *J. Integr. Plant Biol.* 56, 343–349. doi: 10.1111/jipb.12152

Guiry, M. D. (2012). How many species of algae are there? *J. Phycol.* 48, 1057–1063. doi: 10.1111/j.1529-8817.2012.01222.x

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J. B., et al. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol.* 17:148. doi: 10.1186/s13059-016-1012-2

Hamilton, M. L., Warwick, J., Terry, A., Allen, M. J., Napier, J. A., and Sayanova, O. (2015). Towards the industrial production of omega-3 long chain polyunsaturated fatty acids from a genetically modified diatom *Phaeodactylum tricornutum*. *PLoS One* 10:e0144054. doi: 10.1371/journal.pone.0144054

Haslam, R. P., Hamilton, M. L., Economou, C. K., Smith, R., Hassall, K. L., Napier, J. A., et al. (2020). Overexpression of an endogenous type 2 diacylglycerol acyltransferase in the marine diatom *Phaeodactylum tricornutum* enhances lipid production and omega-3 long-chain polyunsaturated fatty acid content. *Biotechnol. Biofuels* 13:87. doi: 10.1186/s13068-020-01726-8

Hempel, F., Lau, J., Klingl, A., and Maier, U. G. (2011). Algae as protein factories: expression of a human antibody and the respective antigen in the diatom *Phaeodactylum tricornutum*. *PLoS One* 6:e28424. doi: 10.1371/journal.pone.0028424

Hsieh-Feng, V., and Yang, Y. (2019). Efficient expression of multiple guide RNAs for CRISPR/Cas genome editing. *aBIOTECH* 1, 123–134. doi: 10.1007/s42994-019-00014-w

Jaussaud, A., Lupette, J., Salvaing, J., Jouhet, J., Bastien, O., Gromova, M., et al. (2020). Stepwise biogenesis of subpopulations of lipid droplets in nitrogen starved *Phaeodactylum tricornutum* cells. *Front. Plant Sci.* 11:1. doi: 10.3389/fpls.2020.00048

Karas, B. J., Diner, R. E., Lefebvre, S. C., McQuaid, J., Phillips, A. P. R., Noddings, C. M., et al. (2015). Designer diatom episomes delivered by bacterial conjugation. *Nat. Commun.* 6:6925. doi: 10.1038/ncomms7925

Kroth, P. G., Bones, A. M., Daboussi, F., Ferrante, M. I., Jaubert, M., Kolot, M., et al. (2018). Genome editing in diatoms: achievements and goals. *Plant Cell Rep.* 37, 1401–1408. doi: 10.1007/s00299-018-2334-1

Leyland, B., Boussiba, S., and Khozin-Goldberg, I. (2020a). A review of diatom lipid droplets. *Biology* 9:38. doi: 10.3390/biology9020038

Leyland, B., Zarka, A., Didi-Cohen, S., Boussiba, S., and Khozin-Goldberg, I. (2020b). High resolution proteome of lipid droplets isolated from the pennate diatom *Phaeodactylum tricornutum* (Bacillariophyceae) strain Pt4 provides mechanistic insights into complex intracellular coordination during nitrogen deprivation. *J. Phycol.* 56, 1642–1663. doi: 10.1111/jpy.13063

Lupette, J., Jaussaud, A., Seddiki, K., Morabito, C., Brugière, S., Schaller, H., et al. (2019). The architecture of lipid droplets in the diatom *Phaeodactylum tricornutum*. *Algal Res.* 38:101415. doi: 10.1016/j.algal.2019.101415

Mikami, M., Toki, S., and Endo, M. (2017). In planta processing of the SpCas9–gRNA complex. *Plant Cell Physiol.* 58, 1857–1867. doi: 10.1093/pcp/pcx154

Moellering, E. R., and Benning, C. (2010). RNA interference silencing of a major lipid droplet protein affects lipid droplet size in *Chlamydomonas reinhardtii*. *Eukaryot. Cell* 9, 97–106. doi: 10.1128/EC.00203-09

Moosburner, M. A., Gholami, P., McCarthy, J. K., Tan, M., Bielinski, V. A., and Allen, A. E. (2020). Multiplexed knockouts in the model diatom Phaeodactylum by episomal delivery of a selectable Cas9. *Front. Microbiol.* 11:5. doi: 10.3389/fmicb.2020.00005

Nymark, M., Sharma, A. K., Sparstad, T., Bones, A. M., and Winge, P. (2016). A CRISPR/Cas9 system adapted for gene editing in marine algae. *Sci. Rep.* 6:24951. doi: 10.1038/srep24951

Pollak, B. (2019). Loop and uLoop assembly. Protocols.io. doi:10.17504/protocols.io.yxnfxme.

Pollak, B., Cerda, A., Delmans, M., Álamos, S., Moyano, T., West, A., et al. (2019). Loop assembly: a simple and open system for recursive fabrication of DNA circuits. *New Phytol.* 222, 628–640. doi: 10.1111/nph.15625

Pollak, B., Matute, T., Nuñez, I., Cerda, A., Lopez, C., Vargas, V., et al. (2020). Universal loop assembly: open, efficient and cross-kingdom DNA fabrication. *Synth. Biol.* 5:ysaa001. doi: 10.1093/synbio/ysaa001

Potapov, V., Ong, J. L., Kucera, R. B., Langhorst, B. W., Bilotti, K., Pryor, J. M., et al. (2018). Comprehensive profiling of four base overhang ligation fidelity by T4 DNA ligase and application to DNA assembly. *ACS Synth. Biol.* 7, 2665–2674. doi: 10.1021/acssynbio.8b00333

Pryor, J. M., Potapov, V., Kucera, R. B., Bilotti, K., Cantor, E. J., and Lohman, G. J. S. (2020). Enabling one-pot Golden Gate assemblies of unprecedented complexity using data-optimized assembly design. *PLoS One* 15:e0238592. doi: 10.1371/journal.pone.0238592

Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* 8, 2281–2308. doi: 10.1038/nprot.2013.143

Sayanova, O., Mimouni, V., Ulmann, L., Morant-Manceau, A., Pasquet, V., Schoefs, B., et al. (2017). Modulation of lipid biosynthesis by stress in diatoms. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 372:20160407. doi: 10.1098/rstb.2016.0407

Serif, M., Dubois, G., Finoux, A. L., Teste, M. A., Jallet, D., and Daboussi, F. (2018). One-step generation of multiple gene knock-outs in the diatom *Phaeodactylum tricornutum* by DNA-free genome editing. *Nat. Commun.* 9:3924. doi: 10.1038/s41467-018-06378-9

Serif, M., Lepetit, B., Weißert, K., Kroth, P. G., and Rio Bartulos, C. (2017). A fast and reliable strategy to generate TALEN-mediated gene knockouts in the diatom *Phaeodactylum tricornutum*. *Algal Res.* 23, 186–195. doi: 10.1016/j.algal.2017.02.005

Sharma, A. K., Nymark, M., Sparstad, T., Bones, A. M., and Winge, P. (2018). Transgene-free genome editing in marine algae by bacterial conjugation – comparison with biolistic CRISPR/Cas9 transformation. *Sci. Rep.* 8:14401. doi: 10.1038/s41598-018-32342-0

Shemesh, Z. (2015). Isolation and characterization of proteins involved in the movement and biogenesis of lipid droplets in the microalga *Phaeodactylum tricornutum*. PhD Thesis. Ben-Gurion Univ. of the Negev, 1–118.

Siloto, R. M. P., Findlay, K., Lopez-Villalobos, A., Yeung, E. C., Nykiforuk, C. L., and Moloney, M. M. (2006). The accumulation of oleosins determines the size of seed oilbodies in *Arabidopsis*. *Plant Cell* 18, 1961–1974. doi: 10.1105/tpc.106.041269

Slattery, S. S., Diamond, A., Wang, H., Therrien, J. A., Lant, J. T., Jazey, T., et al. (2018). An expanded plasmid-based genetic toolbox enables Cas9 genome editing and stable maintenance of synthetic pathways in *Phaeodactylum tricornutum*. *ACS Synth. Biol.* 7, 328–338. doi: 10.1021/acssynbio.7b00191

Stokstad, E. (2020). United States relaxes rules for biotech crops. *Science*. doi: 10.1126/science.370.6521.1145

Strand, T. A., Lale, R., Degnes, K. F., Lando, M., and Valla, S. (2014). A new and improved host-independent plasmid system for RK2-based conjugal transfer. *PLoS One* 9:e90372. doi: 10.1371/journal.pone.0090372

Stukenberg, D., Zauner, S., Dell'Aquila, G., and Maier, U. G. (2018). Optimizing CRISPR/Cas9 for the diatom *Phaeodactylum tricornutum*. *Front. Plant Sci.* 9:740. doi: 10.3389/fpls.2018.00740

Taparia, Y., Zarka, A., Leu, S., Zarivach, R., Boussiba, S., and Khozin-Goldberg, I. (2019). A novel endogenous selection marker for the diatom *Phaeodactylum tricornutum* based on a unique mutation in phytoene desaturase 1. *Sci. Rep.* 9:8217. doi: 10.1038/s41598-019-44710-5

Tran, Q.-G., Yoon, H. R., Cho, K., Lee, S.-J., Crespo, J. L., Ramanan, R., et al. (2019). Dynamic interactions between autophagosomes and lipid droplets in *Chlamydomonas reinhardtii*. *Cell* 8:992. doi: 10.3390/cells8090992

Trentacoste, E. M., Shrestha, R. P., Smith, S. R., Glé, C., Hartmann, A. C., Hildebrand, M., et al. (2013). Metabolic engineering of lipid catabolism increases microalgal lipid accumulation without compromising growth. *Proc. Natl. Acad. Sci. U. S. A.* 110, 19748–19753. doi: 10.1073/pnas.1309299110

Tsai, C. H., Zienkiewicz, K., Amstutz, C. L., Brink, B. G., Warakanont, J., Roston, R., et al. (2015). Dynamics of protein and polar lipid recruitment during lipid droplet assembly in *Chlamydomonas reinhardtii*. *Plant J.* 83, 650–660. doi: 10.1111/tpj.12917

Vancaester, E., Depuydt, T., Osuna-Cruz, C. M., and Vandepoele, K. (2020). Comprehensive and functional analysis of horizontal gene transfer events in diatoms. *Mol. Biol. Evol.* 37, 3243–3257. doi: 10.1093/molbev/msaa182

Wang, M., Mao, Y., Lu, Y., Wang, Z., Tao, X., and Zhu, J. K. (2018). Multiplex gene editing in rice with simplified CRISPR-Cpf1 and CRISPR-Cas9 systems. *J. Integr. Plant Biol.* 60, 626–631. doi: 10.1111/jipb.12667

Willis, A., Chiovitti, A., Dugdale, T. M., and Wetherbee, R. (2013). Characterization of the extracellular matrix of *Phaeodactylum tricornutum* (Bacillariophyceae): structure, composition, and adhesive characteristics. *J. Phycol.* 49, 937–949. doi: 10.1111/jpy.12103

Wood, W. N., Smith, K. D., Ream, J. A., and Kevin Lewis, L. (2017). Enhancing yields of low and single copy number plasmid DNAs from *Escherichia coli* cells. *J. Microbiol. Methods* 133, 46–51. doi: 10.1016/j.mimet.2016.12.016

Xie, K., Minkenberg, B., and Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci.* 112, 3570–3575. doi: 10.1073/pnas.1420294112

Yoneda, K., Yoshida, M., Suzuki, I., and Watanabe, M. M. (2016). Identification of major lipid droplet protein in a marine diatom *Phaeodactylum tricornutum*. *Plant Cell Physiol.* 57, 397–406. doi: 10.1093/pcp/pcv204

Yoneda, K., Yoshida, M., Suzuki, I., and Watanabe, M. M. (2018). Homologous expression of lipid droplet protein-enhanced neutral lipid accumulation in the marine diatom *Phaeodactylum tricornutum*. *J. Appl. Phycol.* 30, 2793–2802. doi: 10.1007/s10811-018-1402-9

Zaslavskaia, L. A., Casey Lippmeier, J., Kroth, P. G., Grossman, A. R., and Apt, K. E. (2000). Transformation of the diatom *Phaeodactylum tricornutum*

(Bacillariophyceae) with a variety of selectable marker and reporter genes. *J. Phycol.* 36, 379–386. doi: 10.1046/j.1529-8817.2000.99164.x

Zienkiewicz, K., and Zienkiewicz, A. (2020). Degradation of lipid droplets in plants and algae—right time, many paths, one goal. *Front. Plant Sci.* 11:579019. doi: 10.3389/fpls.2020.579019

Zienkiewicz, A., Zienkiewicz, K., Poliner, E., Pulman, J. A., Du, Z. Y., Stefano, G., et al. (2020). The microalga Nannochloropsis during transition from quiescence to autotrophy in response to nitrogen availability. *Plant Physiol.* 182, 819–839. doi: 10.1104/pp.19.00854

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 24 June 2022
doi: 10.3389/fpls.2022.941918

Analyses on Flavonoids and Transcriptome Reveals Key *MYB* Gene for Proanthocyanidins Regulation in *Onobrychis Viciifolia*

*Zhongzhiyue Jin, Wenbo Jiang, Yijing Luo, Haijun Huang, Dengxia Yi and Yongzhen Pang**

Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:
Guoyin Kai,
Zhejiang Chinese Medical University, China

Reviewed by:
Chenggang Liu,
University of North Texas, United States
Liangsheng Wang,
Institute of Botany (CAS), China

***Correspondence:**
Yongzhen Pang
pangyongzhen@caas.cn

Specialty section:
This article was submitted to Plant Systems and Synthetic Biology, a section of the journal Frontiers in Plant Science

Received: *12 May 2022*
Accepted: *30 May 2022*
Published: *24 June 2022*

Citation:
Jin Z, Jiang W, Luo Y, Huang H, Yi D and Pang Y (2022) Analyses on Flavonoids and Transcriptome Reveals Key MYB Gene for Proanthocyanidins Regulation in Onobrychis Viciifolia. Front. Plant Sci. 13:941918. doi: 10.3389/fpls.2022.941918

Onobrychis viciifolia (sainfoin) is one of the most high-quality legume forages, which is rich in proanthocyanidins that is beneficial for the health and production of animals. In this study, proanthocyanidins and total flavonoids in leaves of 46 different sainfoin germplasm resources were evaluated, and it showed that soluble proanthocyanidin contents varied greatly in these sainfoin germplasm resources, but total flavonoids did not show significant difference. Transcriptome sequencing with high and low proanthocyanidins sainfoin resulted in the identification of totally 52,926 unigenes in sainfoin, and they were classed into different GOC categories. Among them, 1,608 unigenes were differentially expressed in high and low proanthocyanidins sainfoin samples, including 1,160 genes that were upregulated and 448 genes that were downregulated. Analysis on gene enrichment *via* KEGG annotation revealed that the differentially expressed genes were mainly enriched in the phenylpropanoid biosynthetic pathway and the secondary metabolism pathway. We also analyzed the expression levels of structural genes of the proanthocyanidin/flavonoid pathway in roots, stems, and leaves in the high proanthocyanidin sainfoin *via* RT-qPCR and found that these genes were differentially expressed in these tissues. Among them, the expression levels of *F3'5'H* and *ANR* were higher in leaves than in roots or stems, which is consistent with proanthocyanidins content in these tissues. Among *MYB* genes that were differentially expressed, the expression of *OvMYBPA2* was relatively high in high proanthocyanidin sainfoin. Over-expression level of *OvMYBPA2* in alfalfa hairy roots resulted in decreased anthocyanin content but increased proanthocyanidin content. Our study provided transcriptome information for further functional characterization of proanthocyanidin biosynthesis-related genes in sainfoin and candidate key *MYB* genes for bioengineering of proanthocyanidins in plants.

Keywords: ***Onobrychis viciifolia*, flavonoids, proanthocyanidins, transcriptome, OvMYBPA2**

INTRODUCTION

Onobrychis viciifolia L., also known as sainfoin, is a perennial legume of the *Onobrychis* genus, which is one of the main genus of the Fabaceae family (Carbonero et al., 2011). The centers of genetic diversity of sainfoin originate in southern Central Asia, particularly on the Anatolian plateau of Turkey, the districts of the Caucasus, the margins of the Caspian fringes, parts of Iran, and the mixed swards of Asia Minor (Bhattarai et al., 2016; Mora-Ortiz and Smith, 2018). Sainfoin has a long cultivation history and a wide distribution in temperate and subtropical regions of Europe, North America, Russia, Middle East, and north and northeast Africa (Carbonero et al., 2011; Amirahmadi et al., 2016). Sainfoin, which is abundant in high nutrition, is planted in 23 provinces/autonomous regions of China, such as Gansu, Xinjiang, and Inner Mongolia. Sainfoin has the characteristics of high nutritional value, good palatability, high yield, and strong nitrogen fixation ability, and it is an excellent animal feed resource. It was reported that the ruminant animals fed with sainfoin with moderate level of proanthocyanidins (also called condensed tannins) will reduce the occurrence of bloating and the emission of urinary nitrogen, increase absorption of amino acid (Min et al., 2003) and body weight (Waghorn, 2008), inhibit the damage of pathogens and parasites (Liu et al., 2013; Rivaroli et al., 2019), and promote the ability of nitrogen fixation in environment (Re et al., 2014).

Proanthocyanidins (in short PAs) is one large group of flavonoid compounds, and they were the most widely distributed secondary metabolites and have been widely found in many plant species (Xie and Dixon, 2005). Flavonoids play the important roles with various biological activities for plants, animals, and humans, including but not limited to antiviral, anti-pathogenic bacteria, anti-parasitic, anti-oxidative, anti-inflammatory, anti-allergenic, anti-tumor activities (Pietta, 2000; Kawai et al., 2007; Serafini et al., 2010; Mai et al., 2015; Farhadi et al., 2019; Kopustinskiene et al., 2020; Lalani and Poh, 2020). A variety of flavonoid components are present in plants, including flavonols, flavones, anthocyanins, and proanthocyanidins (Panche et al., 2016), and they are categorized based on the basic skeleton and different oxidation degree (Cheynier et al., 2013). HPLC-MS is an effective tool for the detection of targeted or non-targeted flavonoid compounds in plants (Wu et al., 2013). Based on HPLC-MS analysis, different flavonoid or phenolic compounds were reported in sainfoin. Phenolic compounds in young leaves, young petioles, stems, flower stalks, and flower buds of sainfoin were identified using LC-ESI-MS/MS (Regos et al., 2009). Another study isolated and identified phenolic acid and flavonoids in sainfoin using HPLC-DAD, including arbutin, gallic acid, catechin, epicatechin, epigallocatechin, and procyanidin B2 (Regos and Treutter, 2010). These studies revealed that sainfoin is rich in flavonoids, in particular proanthocyanidins in leaves of some sainfoin, which is rare in legume forages for the prevention of the lethal bloating disease.

Most studies on sainfoin have focused on the identification of phenolic components and the optimization of identification method, as well as evaluation of the bioactivities of related compounds, but the mechanisms regulating flavonoid biosynthetic pathway remain unknown in sainfoin. Biosynthesis pathway of flavonoids has been studied in many model or non-model plant species, including *Arabidopsis thaliana*, *Antirrhinum majus*, *Zea mays*, and *Ginkgo biloba* (Saito et al., 2013; Tohge et al., 2017; Fujino et al., 2018; Su et al., 2020), and these studies identified many biosynthetic genes (e.g., *CHS*, *CHI*, *F3H*, *F3'H*, *F3'5'H*, *DFR*, *ANS*, *ANR*, and *LAR*), as well as different types of transcription factor genes (Liu et al., 2015). Among them, the MBW complex comprising of MYB, bHLH, and WD40 proteins is the major regulatory complex for the regulation of flavonoid and proanthocyanidins pathway (Hichri et al., 2011; Chezem and Clay, 2016). Among the MBW complex, MYBs act as the leading players in the biosynthesis of flavonoids especially proanthocyanidins in seeds or in fruits in many plant species, such as *A. thaliana*, *Vitis vinifera*, *Lotus japonicus*, and *Medicago truncatula* (Baudry et al., 2004; Bogs et al., 2007; Yoshida et al., 2010; Verdier et al., 2012). Therefore, mining and identifying key *MYB* genes that regulate proanthocyanidin biosynthesis in leaves of sainfoin could be beneficial to the bioengineering of proanthocyanidins in other legume forages that lack proanthocyanidins in leave, such as the most important legume forage alfalfa (*Medicago sativa*).

In this study, we analyzed flavonoid and proanthocyanidins content from 46 sainfoin germplasm resources and found proanthocyanidin content varied greatly among them. These special sainfoin germplasm resources with high proanthocyanidins content could be used for breeding. In addition, we selected two of them with relatively high or low proanthocyanidin content for transcriptome analysis and identified a large number of candidate genes potentially involved in proanthocyanidins pathway. Among them, we characterized one MYB transcription gene *OvMYBPA2* and found its over-expression could increase proanthocyanidin content in *M. sativa* hairy roots, by regulating flavonoid flux. Our studies on the function of *OvMYBPA2* provide new clues for the regulation mechanism of proanthocyanidins in sainfoin, as well as new candidate gene for genetic breeding of forages with improved quality.

MATERIALS AND METHODS

Plant Materials

The seeds of 46 sainfoin (*Onobrychis viciifolia*) germplasms (**Supplementary Table 1**) were stored at the Forage Germplasm Bank at the Institute of Animal Science, the Chinese Academy of Agricultural Sciences (IAS-CAAS), Beijing, China. The sainfoin plants were grown from seeds in the CAAS Experimental Station at Lang Fang, Hebei province, China in 2019. The mature leaves were collected during May of 2020 and dried for determination of the contents of proanthocyanidins and total flavonoids.

The seedling of 3-month-old sainfoin plant nos. 25 and 33 was grown in growth chamber (16-h/8-h light/dark, 24°C), and young leaves were collected for transcriptome sequencing and flavonoid profiling. The roots, stems, and leaves of sainfoin nos. 25 and 33 were collected and immediately frozen in liquid nitrogen and then kept at −80°C for RNA extraction and qPCR analyses.

The seeds of alfalfa (*Medicago sativa* L., cultivar Zhongmu no.1) used for the generation of transgenic hairy roots in this study were stored at IAS-CAAS.

Extraction, Quantification, and Identification of Flavonoid Compounds

The extraction and determination of total flavonoids, proanthocyanidins, and anthocyanins were referred to our previous study with slight modifications (Pang et al., 2007, 2009). All reagents were purchased from Sinopharm Chemical Reagent Company (Beijing, China). About 100 mg of dried sainfoin leaves was extracted with 5 ml of 70% acetone (containing 0.5% acetic acid). After ultrasonication at 30°C for 30 min and centrifugation at 2,500 rpm for 10 min, the supernatant was obtained. The supernatant was then transferred into a 15-ml tube, and the residues were re-extracted two times as above. For the determination of soluble proanthocyanidins, 2 ml of supernatant was extracted with 2 ml chloroform to separate hydrophilic from hydrophobic compounds. After centrifugation at 2,500 rpm for 10 min, the upper layer was retained and extracted two times with an equal volume of chloroform. Then, the supernatant was extracted with 500 μl of n-hexane three times. The resulting lower phase was freeze-dried and re-dissolved in 70% acetone extraction solution, and this solution was stored at −20°C for subsequent determination of soluble PAs. The freeze-dried residues were used as a preliminary sample to determine the content of insoluble PAs part.

For the analyses of soluble PAs, 50 μl sample was added into 750 μl DMACA solution (0.2% [w/v] DMACA in methanol-3N HCl [1:1]), and the mixture was determined spectrophotometrically at 640 nm within 15 min. The content of soluble PAs was calculated according to the standard curve with catechin as standard.

For the analyses of insoluble PAs, the insoluble part was dissolved in 5 ml n-butanol-HCl (v/v:95/5). After ultrasonication at room temperature for 30 min and centrifugation at 2,500 rpm for 10 min, the supernatant was detected at 550 nm as A550-1. The supernatant was poured back into the precipitation solution and boiled for 1 h. The absorbance of supernatant was measured again at 550 nm after cool down to room temperature, which was recorded as A550-2. The absorbance difference obtained by subtracting A550-1 from A550-2 was calculated as insoluble PAs, according to the standard curve of procyanidin B1 (Pang et al., 2007, 2009).

For analyses of total flavonoids, the mature leaves of sainfoin or transgenic alfalfa hairy roots were ground in liquid nitrogen and lyophilized at −52°C for 18 h. About 20 mg of samples was extracted with 1 ml of 80% methanol, sonicated for 30 min, and then kept at 4°C for 12 h, following sonication again as above and then centrifuged at 10,000 rpm for 20 min. Afterward, 400 μl water, 30 μl of 5% $NaNO_2$, 30 μl of 10% $AlCl_3$, 200 μl of 1M NaOH, and 240 μl water were added sequentially to 100 μl of the supernatant. The absorptions of mixture of total 1 ml were measured spectrophotometrically at 510 nm. Absorbance values were converted into total flavonoids using a standard curve with quercetin.

TABLE 1 | Content of soluble PAs, insoluble PAs, and total flavonoid determined in leaves of 46 sainfoin germplasm resources.

Sample ID	Soluble PAs (mg/g)	Insoluble PAs (mg/g)	Total flavonoid (mg/g)
1	0.231 ± 0.189	0.520 ± 0.059	0.225 ± 0.009
2	27.319 ± 0.418	0.731 ± 0.082	0.180 ± 0.004
3	24.334 ± 0.837	1.339 ± 0.097	0.184 ± 0.009
4	17.461 ± 1.140	1.005 ± 0.056	0.129 ± 0.001
5	8.626 ± 1.621	0.653 ± 0.051	0.232 ± 0.006
6	18.785 ± 1.332	0.506 ± 0.043	0.127 ± 0.005
7	54.752 ± 1.846	3.412 ± 0.374	0.326 ± 0.010
8	14.343 ± 0.532	0.506 ± 0.043	0.132 ± 0.001
9	30.786 ± 1.766	0.731 ± 0.082	0.207 ± 0.010
10	13.477 ± 0.318	3.975 ± 0.327	0.162 ± 0.004
11	25.646 ± 3.515	0.959 ± 0.046	0.299 ± 0.005
12	39.766 ± 1.611	1.121 ± 0.045	0.267 ± 0.012
13	10.323 ± 1.425	2.157 ± 0.141	0.205 ± 0.010
14	33.675 ± 1.367	1.186 ± 0.030	0.201 ± 0.012
15	12.080 ± 0.454	0.880 ± 0.117	0.176 ± 0.007
16	34.879 ± 1.584	2.790 ± 0.134	0.288 ± 0.013
17	42.691 ± 3.202	1.110 ± 0.074	0.182 ± 0.012
18	22.902 ± 0.683	1.113 ± 0.076	0.213 ± 0.007
19	52.970 ± 0.445	2.249 ± 0.092	0.248 ± 0.009
20	12.730 ± 0.479	2.367 ± 0.146	0.161 ± 0.004
21	0.000	0.618 ± 0.050	0.091 ± 0.005
22	23.046 ± 1.445	0.885 ± 0.022	0.306 ± 0.002
23	48.252 ± 1.643	1.339 ± 0.097	0.392 ± 0.005
24	27.416 ± 1.372	2.249 ± 0.092	0.227 ± 0.005
25	**49.491 ± 2.647**	**0.767 ± 0.052**	**0.277 ± 0.006**
26	15.427 ± 1.096	1.121 ± 0.045	0.208 ± 0.001
27	22.649 ± 1.541	1.113 ± 0.076	0.133 ± 0.002
28	36.179 ± 3.080	1.571 ± 0.142	0.237 ± 0.010
29	33.061 ± 2.455	1.005 ± 0.056	0.193 ± 0.014
30	23.082 ± 0.900	1.467 ± 0.084	0.164 ± 0.006
31	33.964 ± 1.148	1.571 ± 0.142	0.233 ± 0.005
32	55.041 ± 2.497	1.262 ± 0.094	0.294 ± 0.005
33	**0.000**	**0.608 ± 0.044**	**0.241 ± 0.007**
34	3.630 ± 0.280	1.010 ± 0.039	0.199 ± 0.006
35	41.210 ± 3.531	3.082 ± 0.131	0.364 ± 0.013
36	42.450 ± 0.639	2.735 ± 0.045	0.255 ± 0.004
37	18.340 ± 1.018	1.110 ± 0.074	0.249 ± 0.006
38	6.279 ± 1.165	0.579 ± 0.051	0.190 ± 0.003
39	0.000	1.071 ± 0.189	0.121 ± 0.005
40	0.000	1.634 ± 0.103	0.205 ± 0.004
41	0.537 ± 0.388	0.618 ± 0.007	0.228 ± 0.006
42	40.765 ± 2.430	2.690 ± 0.156	0.228 ± 0.006
43	21.674 ± 1.575	4.354 ± 0.298	0.233 ± 0.003
44	6.965 ± 0.353	1.495 ± 0.065	0.173 ± 0.002
45	21.806 ± 0.461	1.262 ± 0.094	0.273 ± 0.011
46	8.590 ± 0.788	1.186 ± 0.030	0.196 ± 0.004

Data are presented as mean ± SD, with biological triplicates. Nos. 25 and 33 that were bold respectively represent sainfoin samples with high or low proanthocyanidins content for subsequent analyses.

The fresh young leaves of sainfoin were grinded and freeze-dried in a vacuum dryer. About 1 ml of 80% MeOH was added into 50 mg leaf sample, sonicated at room temperature for 30 min, and then stored at 4°C overnight. After sonication as above and centrifugation at 13,000 g for 10 min, the supernatants filtered through a 0.22-μm filter were used for subsequent LC-MS analysis. A total of three biological replicates were set up for each sample. For UPLC-MS analyses, chromatographic separation was performed on an Agilent HPLC 6500 system with a hybrid quadrupole time-of-flight (QTOF) Mass Spectrometer (Agilent Technologies, Santa Clara, CA, USA) at 280 nm using an Eclipse XDB-C18 reverse-phase column (4.6 mm ×150 mm, 5 μm). The mobile phases A (100% methanol) and stationary phase B (water containing 1% formic acid) were used at a flow rate of 0.5 ml/min. The parameters and experimental conditions of the instrument were as follows: re-equilibrium with 100% A for 10 min, A:B (v/v) gradient was 5:95 at 0 min, 80:20 at 16 min, 100:0 from 18 −20 min, and 5:95 at 21 min. The column temperature was set at 30°C and the injection volume was 15 μl555. The detection conditions of MS under negative ion (NI) mode were as follows: sample temperature: 350°C, scan range: 100–1,000 (m/z), desolvation gas (N_2) flow: 800 L/h, cone gas flow: 50 L/h, cone voltage−60 V, and capillary voltage 2 kV for NI mode.

For analyses of anthocyanins, 50 mg of fresh hairy roots was extracted with 500 μl of 80% methanol (containing 0.1% hydrochloric acid), sonicated at room temperature for 30 min, centrifuged at 2,500 rpm for 10 min, and transferred the supernatant to a new tube. About 400 μl of supernatant was added into 2-ml centrifuge tube containing 400 μl water and 400 μl chloroform and the mixture was centrifuged as above. About 600 μl of supernatant was detected at 530 nm and the anthocyanin content was calculated as previously reported (Pang et al., 2007, 2009).

RNA Extraction, cDNA Library Construction, High-Throughput Sequencing and Transcriptome *de novo* Assembly and Annotation

Young leaves of 3-month-old sainfoin plant nos. 25 and 33 were collected and used for RNA extraction by using PureLink RNA Mini Kit (Promega, Shanghai, China) according to the manufacturer's instructions. Nanodrop 2000 was used to detect the concentration and purity of RNA (A260/280:1.8-2.2; the concentration: ≥200 ng/μl). About 1% of agarose gel was used to detect RNA integrity, and the RNA integrity value (RIN > 8.0) of those samples was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA). A total amount of 2 μg high-quality RNA per sample was used as input material for the RNA-seq carried out by Zhongxing Bomai Biotechnology Company (Beijing, China).

In details, mRNA purified from total RNAs was transcribed into cDNA using a SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen, Shanghai, China) with reverse transcriptase and random primers. Then, cDNA was combined into the Illumina Novoseq 6000 platform (Illumina, Shanghai, China). Since sainfoin has no reference genome as a control, all clean data were assembled *de novo* through Trinity software (http://trinityrnaseq.sourceforge.net/;version, number: trinityrnaseq-r2013-02-25). The longest sequence of clean reads from contigs was selected as the representative sequence of each gene for the follow-up transcriptome analyses. All the unigene sequences were aligned with the NR, Swissprot, KEGG, and COG databases using BLASTX to obtain the corresponding annotation information. Unigene sequences were deposited in the Sequence Read Archive (SRA) database at NCBI under Bioproject ID: PRJNA810561.

Identification of Differentially Expressed Genes, Function Annotation, and Analysis of DEGs

Gene expression levels were calculated as fragments per kilobase of exon model per million mapped reads (FPKM) using Bowtie and RSEM (http://deweylab.biostat.wisc.edu/rsem/) (Li and Dewey, 2011). DEGs of these two samples were identified from standardized reads using programDeSeq2. The false discovery rate (FDR)-corrected *p*-value (≤0.05) was used as threshold for GO analysis and identification. FDR and fold change (FC) were also used for evaluation of DEGs. Padj < 0.05 and $|\log2(FC)| > 1.5$ were set as the cutoff value to identify DEGs. The heat map was plotted by TB tools (Chen et al., 2020a).

Gene function was annotated based on the following databases: NCBI non-redundant protein sequences (NR), Gene Ontology (GO), eukaryotic Clusters of Orthologous Groups of proteins (KOG), Clusters of Orthologous Groups of proteins (COG), Swiss-Prot (A manually annotated and reviewed protein sequence database), and KEGG Ortholog database. DEGs in KEGG pathways were conducted using FunRich software (version:3.1.3). Clean reads were used to analyze the expression level of the transcripts.

Cloning and Vector Construction

The open reading frame of *OvMYBPA2* was amplified with *KOD* with high fidelity DNA polymerase, and cDNAs obtained from leaves of sainfoin no. 25. The primers OvMYBPA2F and OvMYBPA2R used for cloning of *OvMYBPA2* gene are listed in **Supplementary Table 2**. PCR condition was as follows: 94°C for 3 min, 35 cycles of 94°C for 30 s, 55°C for 30 s, 68°C for 1 min, and followed by a final extension of 68°C for 7 min. PCR products were, respectively, cloned into the Gateway Entry vector pENTR/D-TOPO and verified by sequencing. The resulting entry vector of pENTR/D-OvMYBPA2 was then transferred to the gateway plant transformation destination vector pK7WG2D.1 using the gateway LR reaction to generate pK7WG2D.1-OvMYBPA2 according to the manufacturer's instructions (Invitrogen, Carlsbad, USA).

Analyses of *MYB* Gene Family and Phylogenetic Relationship

The deduced OvMYBPA2 (GenBank accession: OM929200) protein sequences, together with MYB proteins related to flavonoid synthesis from other legume plants

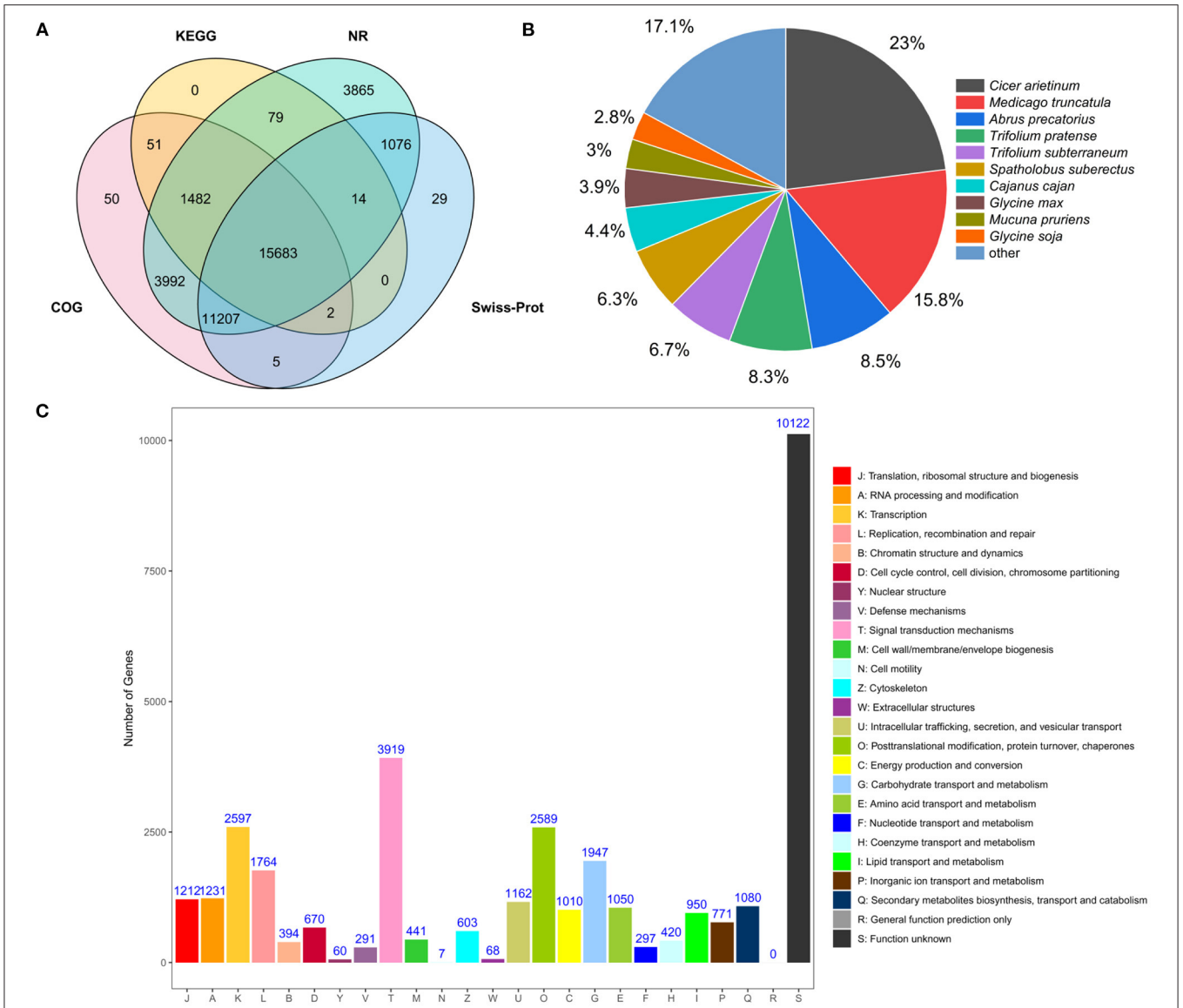


FIGURE 1 | Analyses of transcriptome data from leaves of sainfoin. **(A)** Venn diagram showing the number of unigenes co-annotated to KEGG, NR, COG, and Swiss-Prot database. **(B)** NR annotation shows the distribution of homologous species for unigenes of *O. viciifolia*. Compared to the NR database, the area of the sector represents the proportion of unigenes that are homologous to the corresponding species. **(C)** COG classification of unigenes of *O. viciifolia*. The horizontal axis represents 25 functional categories, and every category was displayed with different color bars, which are denoted by capital letters A~Z. The vertical axis indicates the number of unigenes with corresponding functional categories.

(**Supplementary Table 3**), were used for multiple alignment with DNAMAN 9.0 software. For analyses of phylogenetic relationship, thirty-seven MYB proteins related to flavonoid synthesis from various plant species were aligned with DNAMAN software, and then, the phylogenetic tree was constructed by MEGA 11.0 software using the neighbor-joining (NJ) model with 1,000 bootstrap replications.

Analyses of Gene Expression by Real-Time Quantitative PCR

Samples from roots, stems, and leaves of sainfoin or alfalfa hairy roots were ground in liquid nitrogen, and total RNAs were extracted from different tissues using Eastep® Super total RNA Extraction Kit (Promega, Shanghai, China) according to the manufacturer's instructions. The first-stand cDNA was synthesized from 500 ng total RNAs by using FastKing gDNA Dispelling RT SuperMix (Tiangen, Beijing, China) according to the manufacturer's protocols. The RT-qPCRs were carried out on ABI 7500 real-time Detection System (Applied Biosystems, USA) using a 2×RealStar Green Fast Mixture (GeneStar, Being, China). Actin-related protein 4A genes were used as house-keeping gene as control in the RT-qPCR. The RT-qPCR conditions were as follows: 95°C for 2 min, 40 cycles of 95°C for 15s and 60°C for 34s, 95°C for 15s, 60°C for 1 min, 95°C for 15s, and 60°C for

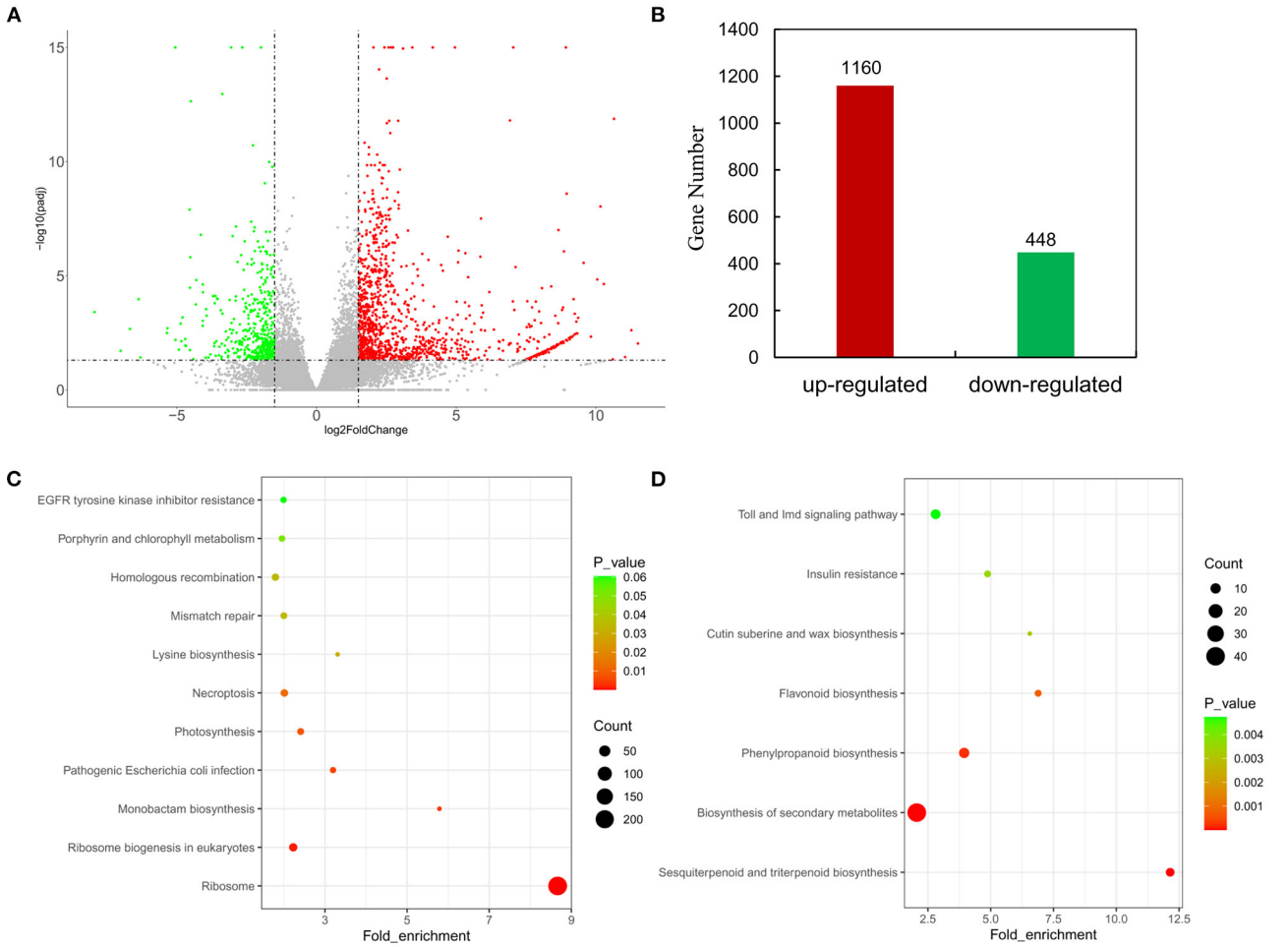


FIGURE 2 | Analyses on DEGs in leaves of two *O. viciifolia* samples. **(A)** Volcano plots of DEGs. The horizontal axis represents the values of Log_2FoldChange of DEGs in the samples of high PAs and low PAs groups, and the vertical axis represents the statistical significance of differences with *p*-value. Red represents upregulated DEGs and green represents downregulated DEGs. **(B)** Gene number of upregulated DEGs and downregulated DEGs. **(C)** Enrichment analyses of KEGG function of upregulated DEGs with 1,160 genes. **(D)** Enrichment analyses of KEGG function of downregulated DEGs with 448 genes. The horizontal axis is the fold enrichment factor, and the vertical axis is the KEGG pathway. The bubbles represent the number of genes. Different colors represent significance of differences (*p*-value).

15s. Each reaction was performed with three biological replicates, and the relative transcript levels were calculated compared to the internal control using the $2^{-\Delta\Delta CT}$ method. The data were presented as means ± standard. The primer used in this study was designed using the Premier 5.0 software, and primer sequences are listed in **Supplementary Table 2**.

Generation and Identification of Transgenic Alfalfa Hairy Roots Mediated by *Agrobacterium*

The pK7WG2D.1-OvMYBPA2 plasmids confirmed by sequencing were transformed into the *Agrobacterium rhizogenes* strain Arqual I. The seeds of alfalfa were treated in sequence with concentrated sulfuric acid, sodium hypochlorite, and sterile water, spread flat on MS medium, and vernalized in a refrigerator at 4°C for 3 days. The hypocotyl parts were infected with Arqual I containing the above-mentioned plasmids and incubated on the F agar medium until the generation of hairy roots. The hairy roots were then transferred onto B5 medium containing 50 mg/L kanamycin for selection. Hairy roots of about 90 days old were ground in liquid nitrogen, and DNA was extracted with modified CTAB method (Drabkova, 2014). PCRs for the identification of positive hairy root lines were performed with *Taq* DNA polymerase with the same above PCR condition for gene cloning. The transgenic hairy roots of 90 days old were confirmed by PCR and RT-qPCR, and the positive lines were selected for analysis of anthocyanins, soluble PAs, and total flavonoids using the above-mentioned methods.

RESULTS

Analyses of Proanthocyanidins and Total Flavonoids in Leaves of Sainfoin

During the quality evaluation of sainfoin, main flavonoids including soluble PAs, insoluble Pas, and total flavonoids were determined spectrophotometrically in 46 germplasm resources

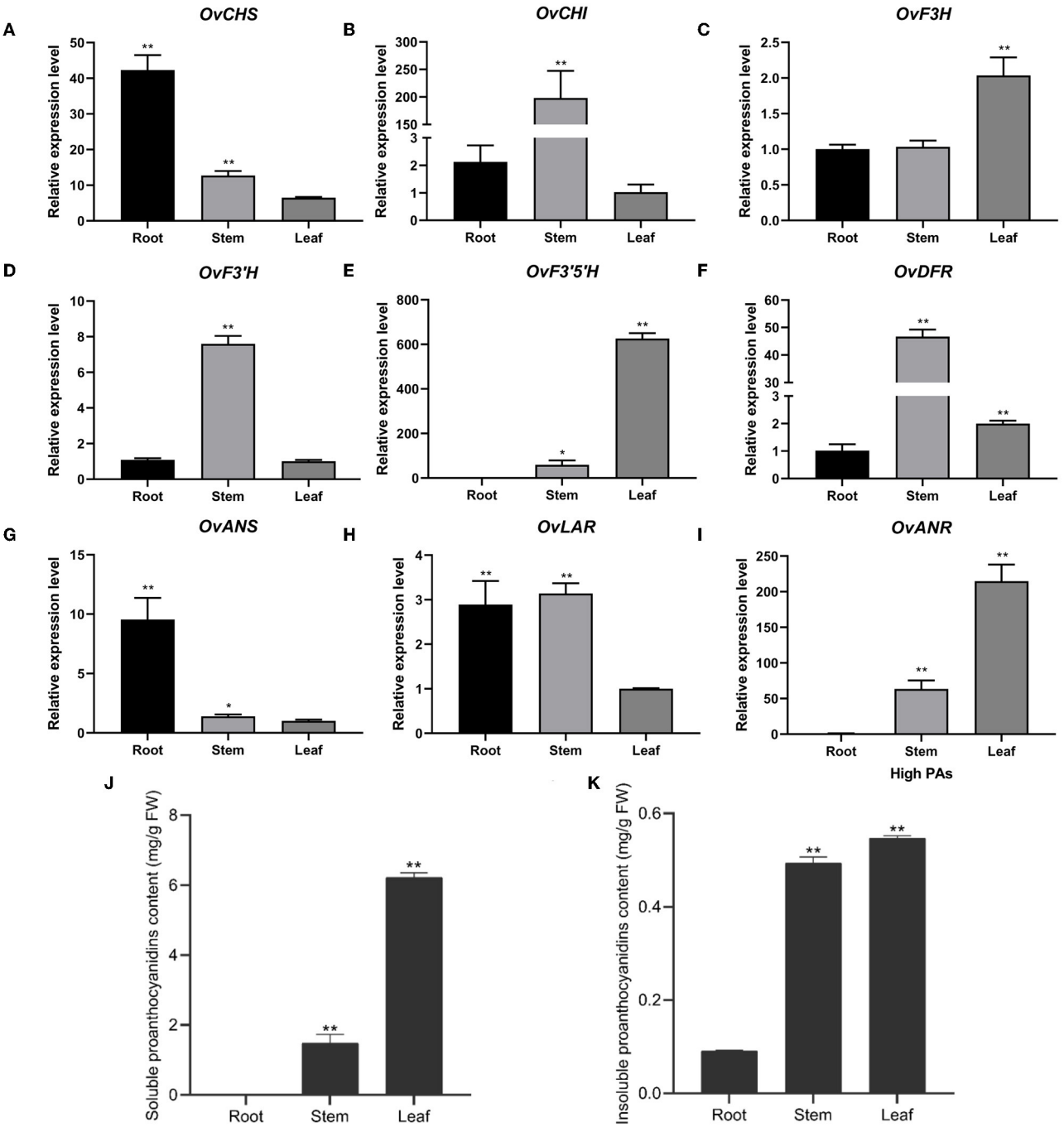


FIGURE 3 | Relative expression levels of pathway genes in different tissues of *O. viciifolia.* The horizontal axis is different tissues (root, stem, and leaf) of HPAs sainfoin, and the vertical axis is relative expression levels. *OvCHS* **(A)**, *OvCHI* **(B)**, *OvF3H* **(C)**, *OvF3'H* **(D)**, *OvF3'5H* **(E)**, *OvDFR* **(F)**, *OvANS* **(G)**, *OvLAR* **(H)**, and *OvANR* **(I)**. **(J,K)** Soluble **(J)** and insoluble proanthocyanidins **(K)** content in root, stem, and leaf of HPAs sainfoin. Error bar depicts the standard error of mean ± SD of three biological replicates. Significance of differences was represented with one asterisk ($p < 0.05$) and two asterisks ($p < 0.01$).

collected from home and abroad (**Supplementary Table 1**). It was revealed that soluble PA contents in leaves of 46 sainfoin germplasm resources ranged from 0 to 55.041 mg/g (**Table 1**). Among them, soluble PA contents in five of them were more than 45 mg/g (nos. 7, 19, 23, 25, and 32) (**Table 1**), whereas no PAs were detected in another four of them (nos. 21, 33, 39, and 40) (**Table 1**). Measurement on insoluble PAs content showed that it ranged from 0.506 to 4.354 mg/g, and the insoluble PAs did not

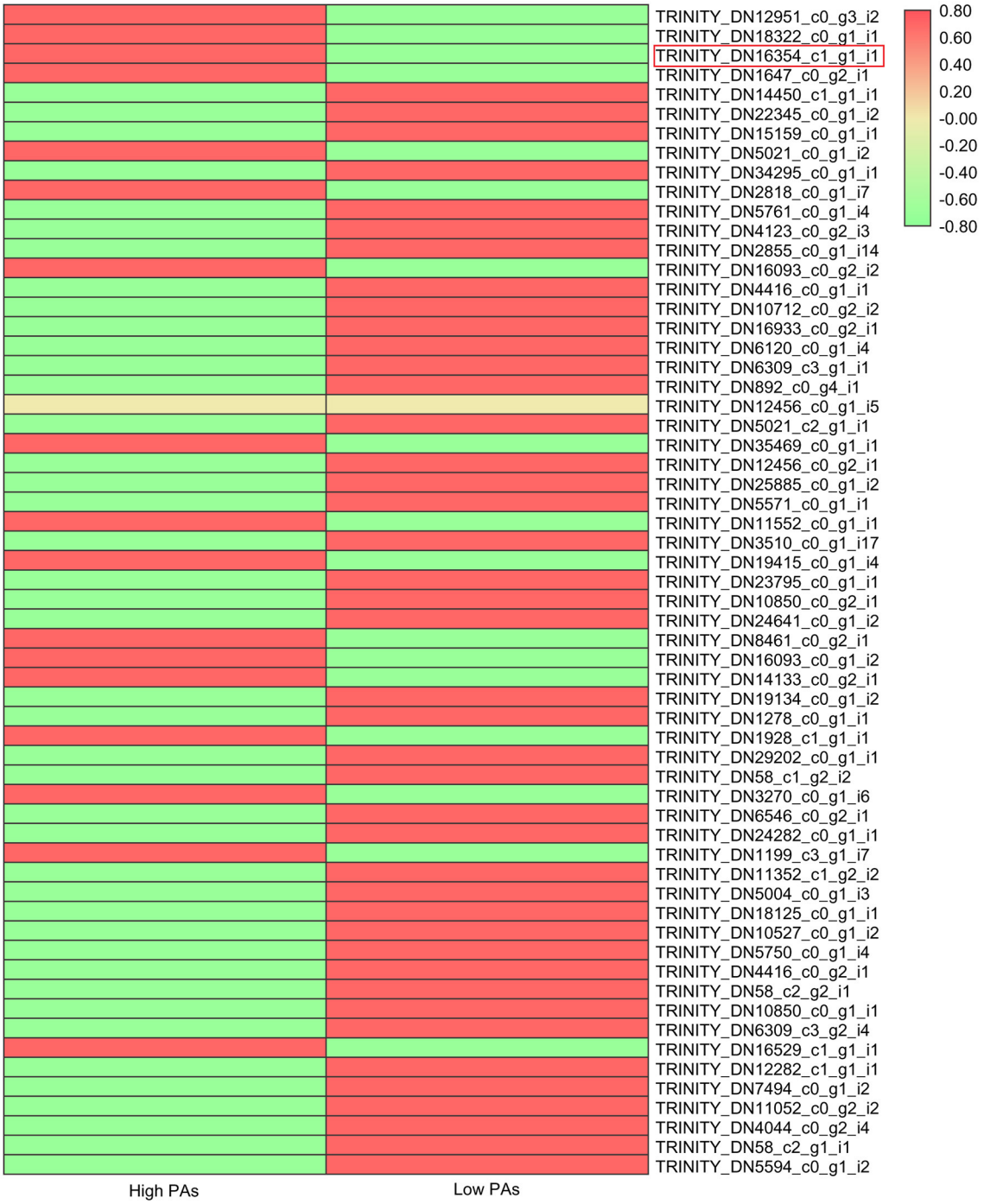


FIGURE 4 | Expression levels of MYB transcription factor genes in two sainfoin samples. The heat map was drawn by TBtools with the Log_2FPKM values of 60 MYBs in *O. viciifolia* with high PAs and low PAs. Different colors depict different expression levels, red means higher expression levels, and green means lower expression levels. Red box represents the MYB named as *OvMYBPA2* that was subsequently cloned and functionally validated in this study.

show significant difference as soluble PAs among 46 germplasm resources (**Table 1**). We also determined total flavonoids and found that it ranged from 0.091 to 0.396 mg/g (**Table 1**). These data showed that soluble PAs content varied significantly in leaves of these 46 sainfoin germplasm resources, whereas insoluble PAs content and total flavonoid content did not (**Table 1**).

A total of four sainfoin germplasm resources with relative higher soluble PAs content (nos. 7, 19, 25, and 32), and another

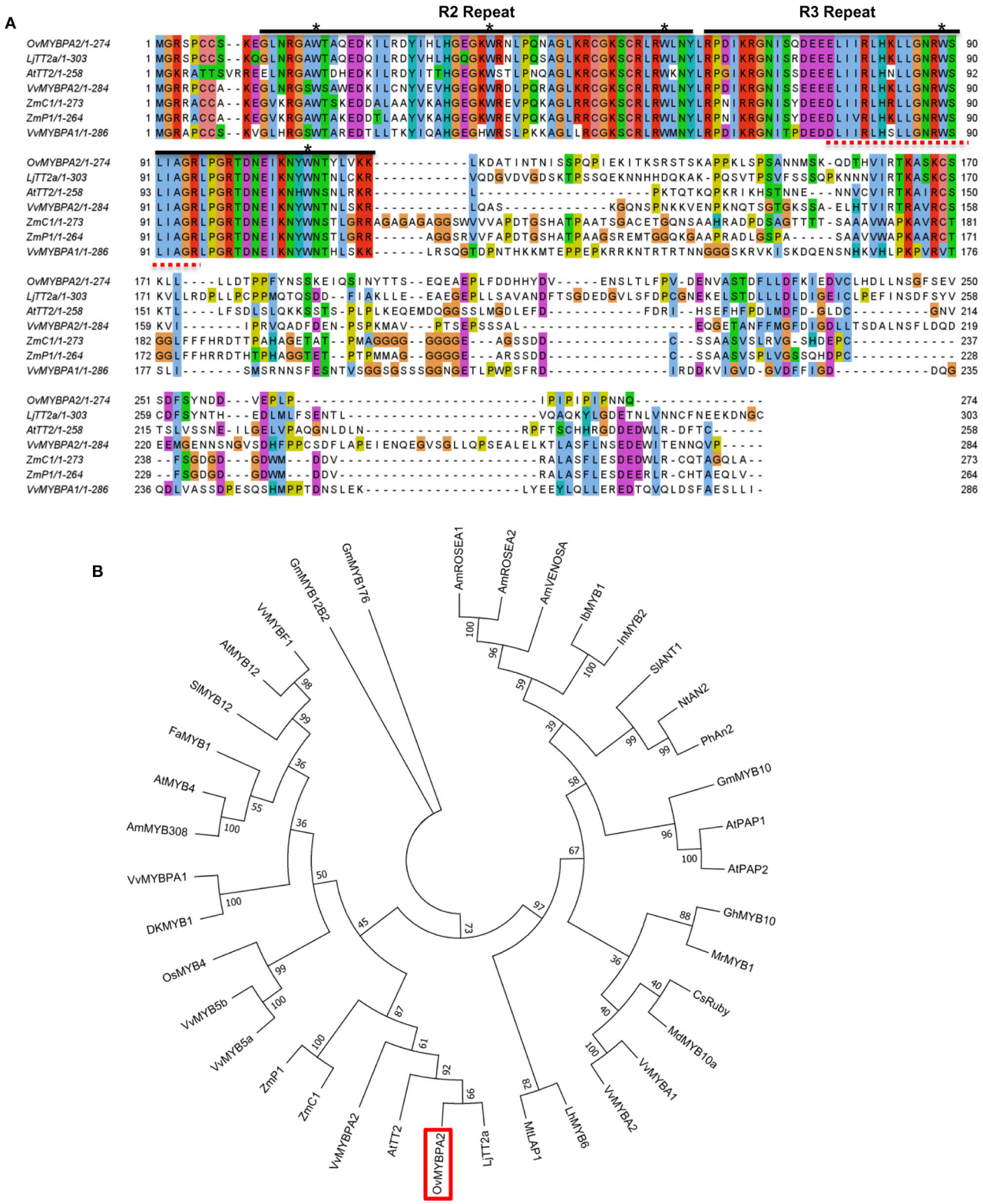


FIGURE 5 | Multiple sequence alignment and evolutionary relationship of OvMYBPA2. **(A)** Alignment of amino acids sequence of *OvMYBPA2* with other MYBs (e.g., LjTT2a, AtTT2, VvMYBPA1, VvMYBPA2, ZmC1, and ZmP1) that were related to proanthocyanidins biosynthesis. The R2 and R3 repeats of MYB conserved domain were marked with black line. The five asterisks indicate conserved W residues. The bHLH-binding domains were indicated in dashed red lines **(B)** phylogenetic analysis of OvMYBPA2 and other MYB proteins related to proanthocyanidins, anthocyanidins, and flavonoid biosynthesis. OvMYBPA2 was highlighted with a red box. The sequence information for all the MYB transcription factors was downloaded from the GenBank database and listed in **Supplementary Table 3**.

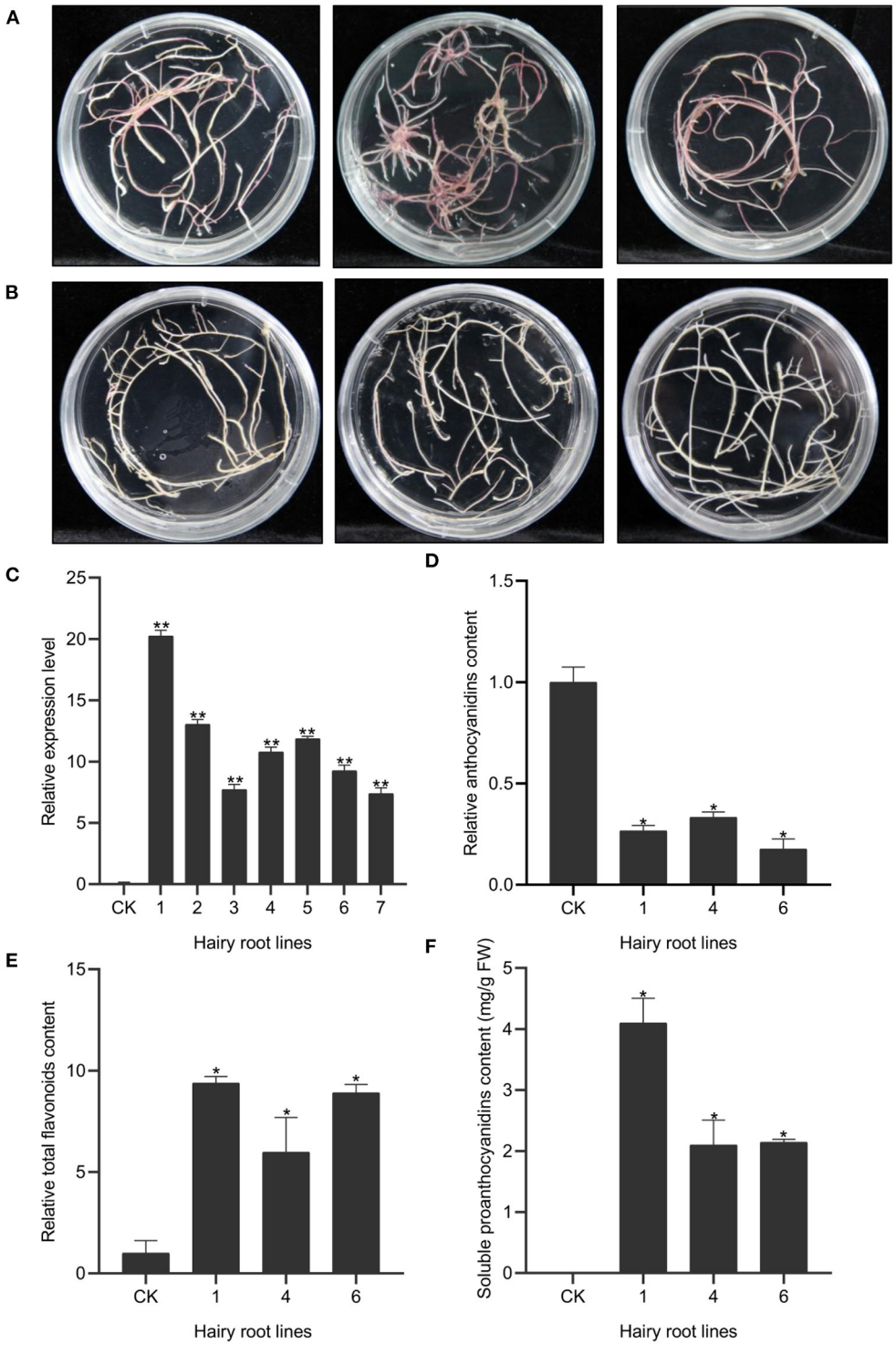


FIGURE 6 | Over-expression of *OvMYBPA2* in alfalfa hairy roots. **(A)** Phenotype of control hairy roots grown on B5 medium. **(B)** Phenotype of *OvMYBPA2*-over-expressing hairy roots grown on B5 medium. **(C)** Relative expression levels of *OvMYBPA2* in transgenic hairy roots and control hairy roots (CK). **(D)** Proanthocyanidins content in different transgenic hairy root lines. **(E,F)** Relative content of anthocyanidin **(E)** and flavonoids **(F)**. Error bar depicts the standard error of mean ± SD of three biological replicates. Significance of differences are indicated with one asterisk ($p < 0.05$) and two asterisk ($p < 0.01$).

four (nos. 21, 33, 39, and 40) with relatively lower soluble PAs content, were selected and their leaves were tested at different sampling times. Among them, the soluble PA content in nos. 25 and 33 maintained stable, and they thus were selected for further analyses as high PA sample (no. 25, HPAs) and low PA sample (no. 33, LPAs).

The composition of flavonoids in leaves of sainfoin nos. 25 and 33 was further subjected for LC-MS analyses. Several peaks were identified and their content in these two samples was different (**Supplementary Table 4**; **Supplementary Figure 1**). Combining UV spectrograms, mass spectra, and previous reports (Regos et al., 2009; Regos and Treutter, 2010), the contents of eight compounds, namely, quercetin-3-*O*-rutinoside, kaempferol-3-*O*-rutinoside, kaempferol, quercetin, L-tryptophan, *p*-coumaroylquinic acid, myricetin 3-*O*-rhamnoglucoside, and caffeoylquinic acid, were different in these two samples (**Supplementary Table 4**). Among them, four compounds, including quercetin-3-*O*-rutinoside, kaempferol-3-*O*-rutinoside, quercetin, and kaempferol, were higher in no. 25 than in no. 33, whereas quercetin, L-tryptophan, myricetin 3-*O*-rhamnoglucoside, and caffeoylquinic acid were only found in no. 25. A total of five remaining peaks could not be identified based on available data, and their contents in nos. 25 and 33 were different (**Supplementary Table 4**). Meanwhile, no monomers, dimers, or other low molecular oligomers were identified in leaves of these two sainfoin samples by LC-MS under present detection condition (**Supplementary Figure 1**).

Transcriptome Sequencing and Annotation

In this study, Illumina Hi-Seq paired-end sequencing technology was used to analyze the transcriptome of sainfoin leaves. Total RNAs were extracted and assessed, and no. 25 with high PAs and no. 33 with low PAs were established and sequenced with triplicates. As a result, a total of more than 40 billion high-quality clean reads after filtration were produced. The Q20 and Q30 of each sample were not less than 98.3 and 94.5%, respectively, and the GC content of each samples was between 43 and 45% (**Supplementary Table 5**). The clean reads were assembled by Trinity software to 52,926 transcripts with an average length of 1,063 bp and an N50 contig size of 1,587 bp after removing redundant transcripts, with the GC content of assembled unigenes of 40.14% (**Supplementary Table 5**). The length of the largest transcript and smallest transcript was, respectively, 14,543 and 301 bp. The length of N50 was higher than the average length of assembled unigenes (**Supplementary Table 6**). These data indicated that the quality of this sequencing met the requirements for subsequent transcriptome analysis.

All unigenes in accordance with the public databases NR, COG, KOG, KEGG, GO, Swisss-Prot were searched using BLASTX (E-value$<10^{-5}$) for functional annotations. A total of 52,926 unigenes (100%) were annotated in the six databases. A total of 37,398 transcripts (70.6%) were significantly matched with NR database, and 32,472 (61.4%), 18,743 (35.4%), 17,312 (32.7%), 17,912 (33.8%), and 28,016 (52.9%) unigenes matched with COG, KOG, KEGG, GO, Swiss-Prot databases, respectively (**Supplementary Table 7**). It can be found that some unigenes could not be annotated with these database, and it is possible that these unannotated unigenes are unique to sainfoin and require more advanced sequencing techniques for further analyses. A Venn diagram showed that a total of 15,683 unigenes were co-annotated in KEGG, NR, COG, and Swiss-Prot databases (**Figure 1A**).

By comparing with the NR library, the annotated transcript of sainfoin was related to the following species with different transcriptome amount and percentage: *Cicer arietinum* (8,609, 23%), *Medicago truncatula* (5,913, 15.8%), *Abrus precatorius* (3,194,8.5%), *Trifolium pratense* (3,101, 8.3%), *Trifolium subterraneum* (2,515,6.7%), and *Spatholobus suberectus* (2,369, 6.3%) (**Figure 1B**). The results revealed that sainfoin had a relatively high homology with *C. arietinum* and *M. truncatula*, and they both belong to the legume family as sainfoin.

Using the GO database, clustering analysis of gene functions was annotated and classified according to the three aspects of biological process (BP), cellular component (CC), and molecular function (MF), and these three terms were mainly categorized into 60 subcategories. For the BP category, organic substance metabolic process (10,167), cellular metabolic process (10,142), primary metabolic process (9,514), and nitrogen compound metabolic process (8,558) were the most dominant subcategories (**Supplementary Figure 2**). In the subcategories of cellular components, genes were mainly annotated in organelle (11,620), cytoplasm (9,719), membrane (5,975), and cell periphery (4,136). The most enriched terms in molecular function were transferase activity (3,599), organic cyclic compound binding (3,570), heterocyclic compound binding (3,553), and hydrolase activity (3,526, **Supplementary Figure 2**).

For COG annotation, a total of 32,472 unigenes were divided to 25 groups. Signal transduction mechanism (T, 3,919) was the most abundant group (**Figure 1C**). According to the classification results, 1,080 genes were annotated to the group of secondary metabolites biosynthesis, transport, and catabolism (Q), and it can be hypothesized that these genes may be closely associated with metabolome variability in the leaves of sainfoin (**Figure 1C**).

Analysis of KEGG pathway genes was also performed, and a total of 17,312 unigenes (32.7%) were mapped to KEGG database. These unigenes were divided into five groups, namely, metabolism, genetic information processing, environmental information processing, cellular processes, and organismal systems, and they were further divided into 32 pathways (**Supplementary Figure 3**). Among these, carbohydrate metabolism (1,360), translation (1,147), signal transduction (2,682), transport and catabolism (1,110), and immune system (1,074) were the top five subcategories. Among these pathways, 124 and 17 genes were, respectively, associated with phenylpropanoid biosynthesis and flavonoid biosynthesis, and ko00940 was the most abundant pathway with enriched genes in phenylpropanoid biosynthesis (**Supplementary Figure 3**).

Analyses of Differentially Expressed Genes Between HPAs and LPAs Sainfoin

To investigate DEGs that may be related to flavonoid metabolism between HPAs and LPAs, especially those related to PA biosynthesis, we performed a comparative analysis of the

transcriptome of these two samples. The intergroup correlation values were not less than 0.93 among these two samples with triplicates (**Supplementary Table 8**), indicating that these six samples have high correlation and the subsequent analysis on differentially expressed genes was reliable.

The distribution of differences in gene expression levels was determined between HPAs and LPAs by constructing a volcano plot (**Figure 2A**). |Log2(fold change)| $\geq$1.5 and Padj$\leq$0.05 were set as thresholds to screen significant differentially expressed genes, and it was found that 1,608 genes were differentially expressed between the transcriptome of HPAs and LPAs, including 1,160 genes (expression level in HPAs/LPAs) that were upregulated and 448 genes that were downregulated (**Figure 2B**).

To analyze the major functional terms of these DEGs, we performed GO enrichment analysis with both upregulated and downregulated genes, respectively. Protein-containing complex, gene expression, ribonucleoprotein complex, structural molecule activity, and ribosome were the top five terms of GO enrichment analysis for the upregulated genes (**Supplementary Figure 4**). The top four terms of GO enrichment analysis for the downregulated genes were anion transmembrane transport, anion transmembrane transporter activity, organic anion transport, and anion transport (**Supplementary Figure 5**).

To analyze potential pathways involved in flavonoid metabolism, KEGG enrichment analysis was performed with all DEGs. KEGG enrichment analysis with the fold enrich factor, Q-value, and the count of genes in a given pathway were shown in bubble plots (**Figures 2C,D**). A total of 1,160 upregulated genes were mapped to 244 pathways and mainly enriched in ribosome (218), ribosome biogenesis in eukaryotes (18), pathogenic Escherichia coli infection (7), photosynthesis (10), and necroptosis (13, **Figure 2C**). All 448 downregulated genes were mapped to 132 pathways and mainly annotated to biosynthesis of secondary metabolites (40), sesquiterpenoid and triterpenoid biosynthesis (7), phenylpropanoid biosynthesis (10), and flavonoid biosynthesis (5, **Figure 2D**). The KEGG analysis for the downregulated genes showed that HPAs sainfoin had more DEGs enriched in secondary metabolite biosynthesis pathway and phenylpropanoid biosynthesis pathway than in flavonoid synthesis pathway compared with LPAs sainfoin. This result suggested that the PAs content in the leaves of HPAs sainfoin might be due to the downregulated of DEGs in the flavonoid synthesis pathway.

Analysis of Differentially Expressed Genes Associated With PA Pathway

To understand the mechanism of flavonoid synthesis in sainfoin, we analyzed and categorized genes involved in flavonoid pathway from the transcriptome data and showed the expression levels of several representative structural genes in HPAs and LPAs sainfoin. Gene expression patterns of structural gene including *CHS*, *CHI*, *F3H*, *F3'H*, *F3'5'H*, *DFR*, *ANS*, *LAR*, and *ANR* were represented by the FPKM values and Log_2FC (**Supplementary Table 9**). According to these data, it was obviously that the structural genes related to flavonoids biosynthesis had different expression patterns between HPAs and LPAs (**Supplementary Table 9**). It was found that upstream genes of the flavonoid pathway were mostly downregulated, such as *CHS* (TRINITY_DN1298_c0_g1_i1, TRINITY_DN613_c0_g1_i2, TRINITY_DN20751_c0_g1_i1, TRINITY_DN320_c0_g1_i4), *CHI* (TRINITY_DN12115_c0_g1_i1, TRINITY_DN586_c0_g2_i3), *F3H* (TRINITY_DN4402_c0_g1_i1), *F3'H* (TRINITY_DN1854_c0_g1_i3), and *F3'5'H* (TRINITY_DN2008_c0_g1_i3) (**Supplementary Table 9**). However, several downstream genes were upregulated genes such as *DFR* (TRINITY_DN282_c0_g1_i15), *LAR* (TRINITY_DN11234_c1_g1_i2), and *ANR* (TRINITY_DN293_c0_g2_i1) in the later flavonoid pathway for PA biosynthesis (**Supplementary Table 9**).

To further investigate the expression of nine structural genes in different tissues, we analyzed their expression level using RT-qPCR in the roots, stems, and leaves of sainfoin with HPAs. It was shown that the expression of *CHS* and *ANS* shared a similar expression pattern which were both higher in roots than in stems and leaves (**Figures 3A,G**). *CHI*, *F3'H*, *DFR*, and *LAR* were highly expressed in stems than in roots and leaves (**Figures 3B,D,F,H**), and the expression levels of *F3H*, *F3'5'H*, and *ANR* were higher in leaves than in stems or roots (**Figures 3C,E,I**). It was worth noting that the expression levels of *F3'5'H* and *ANR* were relatively higher than other genes, in particular in leaves (**Figures 3E,I**). In addition, we also analyzed both soluble PAs and insoluble PAs in roots, stems, and leaves of in this sainfoin and found that both soluble and insoluble PA contents were higher in leaves than in roots or stems (**Figures 3J,K**). It was thus speculated that *F3'5'H* and *ANR* are the key genes for PAs biosynthesis in leaves of sainfoin.

Many studies have shown that flavonoid pathway was mainly regulated by transcription factor complex composing of MYB, bHLH, and WD40. Hence, we also investigated the expression levels of all putative MYB, bHLH, and WD40 genes that were differentially expressed in two sainfoin samples (**Figure 4**; **Supplementary Figures 6, 7**). The expression levels of total 60 *MYBs*, 113 *bHLHs*, and 19 *WD40* genes were compared and shown in heat map (**Figure 4**; **Supplementary Figures 6, 7**). Among the MBW complex proteins, MYB is the key player by controlling the biosynthesis of PAs *via* enhancing or inhibiting the expression of structural genes (Hichri et al., 2011); therefore, we focused on *MYB* genes and found that 17 *MYB* genes were upregulated in HPAs when compared to LPAs (**Figure 4**), suggesting that these TFs may be associated with PAs biosynthesis in sainfoin. Among them, one unigene (TRINITY_DN16354_c1_g1_i1) was upregulated by 1.5-fold in the HPAs sample (**Figure 4**), and it showed high sequence similarity with other *MYB* genes involved in PA pathway, which were therefore selected as candidate genes for further analyses.

Multiple Sequence Alignment and Phylogenetic Analysis of *OvMYBPA2* Gene

To further characterize the function of the selected *MYB* gene, its open reading frame was cloned and named as *OvMYBPA2* (under GenBank accession number OM929200). The ORF of

OvMYBPA2 is 825 bp in length, which encodes a putative protein of 274 amino acid. The deduced OvMYBPA2 protein showed ∼40% similarity to ZmC1 from maize, VvMYBPA1 and VvMYBPA2 from grape, AtTT2 from *Arabidopsis*, and 53% to LjTT2a from *L. japonicus* (**Figure 5A**). Multiple sequence alignment with *MYB* genes from other plant species revealed that OvMYBPA2 has a highly homologous conserved R2R3 DNA-binding domains at the N-terminus, whereas the C-terminus differed significantly in sequence and length (**Figure 5A**), and the bHLH-binding domain was identical to those of LjTT2a, VvMYBPA1/A2, ZmC1, and ZmP1 (**Figure 5A**).

OvMYBPA2 and several MYBs from other plant species that regulate PAs, anthocyanins, and flavonoids biosynthesis were used for phylogenetic analysis, and it showed that these MYBs formed three distinct clades, including the anthocyanin clade containing AtPAP1, AtPAP2, and MtLAP1, the PA clade containing AtTT2, and the flavonoid clade containing AtMYB5 and AtMYB4 (**Figure 5B**). OvMYBPA2 was clustered within the PA clade containing LjTT2a, AtTT2, VvMYBPA2, ZmC1, and ZmP1, suggesting that OvMYBPA2 may have a similar function in PAs regulation in sainfoin (**Figure 5B**).

Over-expression of *OvMYBPA2* Gene in Alfalfa Hairy Roots

To further determine the *in vivo* regulatory function of OvMYBPA2, we over-expressed it in hairy roots of alfalfa, a target forage crop for PA bioengineering. A total of fifty-three positive transgenic hairy root lines were successfully generated and further confirmed by PCRs (**Figures 6A,B**). The expression level of *OvMYBPA2* was confirmed by RT-qPCR analyses in seven representative hairy root lines, but no *OvMYBPA2* expression was detected in the vector control line (**Figure 6C**). RT-qPCRs showed that the expression of *OvMYBPA2* was activated to various levels in different hairy roots lines as compared to the control hairy roots (**Figure 6C**). Among them, three lines (nos. 1, 4, and 6) with relatively high expression level and stable growth condition were propagated and selected for further analyses.

We found that anthocyanin contents in hairy root lines nos.1, 4, and 6 were significantly reduced by 73, 67, and 82% as compared to the control line (**Figure 6D**). Soluble PA content was 4.10, 2.10, and 2.14 mg/g FW in the hairy root line nos.1, 4, and 6, respectively, and that in the control line was negligible (**Figure 6E**). Total flavonoids in hairy root lines nos.1, 4, and 6 were 9.38, 5.98, and 8.90 times as compared to the control hairy root line (**Figure 6F**). Taken together, these data clearly indicated that over-expression of *OvMYBPA2* promoted flavonoid accumulation, in particular proanthocyanidins accumulation in alfalfa hairy roots.

DISCUSSION

Sainfoin is a high-quality forage grass, which is widely cultivated in western China, and it is the second legume grass after alfalfa in term of cultivation area and yield. In particular, some sainfoin contained a large amount of flavonoid, specifically PAs, which is absent in foliage of alfalfa. In this study, we analyzed flavonoid and PAs content in leaves of 46 sainfoin germplasm resources and found that soluble PA varied significantly, but total flavonoids and insoluble PAs did not show much difference (**Table 1**), and these data indicated that the difference in flavonoids content is mainly contributed by soluble PAs in sainfoin. This characteristic of sainfoin is distinct from the other forage crop alfalfa and its close relative *M. truncatula*, which did not accumulate PAs in the foliage tissues, but in the seed coats (Pang et al., 2007).

In addition to soluble PAs, a few flavonoid compounds were also identified, although they showed difference in composition, but their amount was not as significant as soluble proanthocyanidins (**Table 1**; **Supplementary Figure 1**). It was found that eight flavonoid compounds accumulated in both sainfoin, which was consistent with a previous report in terms of flavonoid composition (Regos and Treutter, 2010). It was notable that no proanthocyanidin monomers such as (epi)catehicn, (epi)afzenin or (epi)gallocatechins, or dimmers were detected in leaves of sainfoin by HPLC-MS in our study (**Supplementary Figure 1**), which was different from other plant, for examples, *Arabidopsis*, *M. truncatula,* or grape that accumulate free monomer of different types (Lepiniec et al., 2006; Pang et al., 2007; Chen et al., 2020b). On the one hand, these phenomena could be explained by extraction and identification method for these two representative samples. On the other hand, it is possible that a large amount of proanthocyanidin were in the form of high polymerization degree, and the identification of high molecule PAs is still challenging. Therefore, the degree of polymerization of proanthocyanidins in sainfoin requires further investigation with improved method in the near future.

With the development of molecular biology technologies, RNA-sequencing has become an effective strategy to investigate the transcriptome of non-model plant species. This method has been widely used to mine and investigate genes related to a certain pathway in plants. For example, Bai et al. analyzed the key genes related to flavonoid synthesis in *Scutellaria viscidula* using transcriptome sequencing (Bai et al., 2018), and Sun et al. mined *MYB* genes regulating flavonoid synthesis in *Zanthoxylum bungeanum* with transcriptome analysis (Sun et al., 2019). We also identified a number of genes related to isoflavonoid puerarin biosynthesis in kudzu in the previous study (Shen et al., 2021). Currently, no genome information is available for sainfoin, so this study is the first report on transcriptome sequencing of sainfoin to mine genes regulating biosynthesis of proanthocyanidins in sainfoin. A total of two sainfoins with high and low proanthocyanidin contents were used to compare the expression level of potential pathway genes, and our results showed that the transcripts of sainfoin leaves were closely related to legume plants such as *C. arietinum*, *M. truncatula*, and *T. pratense* (**Figure 1**), indicating that they may share common functional genes as they share close evolutionary relationship (Koenen et al., 2020).

The comparison of these two sainfoin resulted in the identification of more than one thousands unigenes that were

differentially expressed in these two samples (**Figure 2**), with enriched genes in phenylpropanoid pathway and secondary metabolic pathway, and these genes are most like involved in the regulation of proanthocyanidins. Analysis of these differentially expressed genes showed that a number of *MYB*, *bHLH*, and *WD40* genes were differentially expressed (**Figure 4**; **Supplementary Figures 6, 7**). In particular, RT-qPCR showed that the key enzyme genes of flavonoid synthesis pathway were expressed in different tissues at different levels (**Figure 3**). In particular, *F3'5'H* and *ANR* genes were preferentially expressed in leaves of HPAs sainfoin, which is likely responsible for high content of PAs in leaves (**Figure 3**). *F3'5'H* is responsible for the hydroxylation of flavonoid in 3' and 5' hydroxy group in the B-ring, which is likely responsible for epigallocatechin and further for proanthocyanidin biosynthesis in sainfoin as previously reported (Regos and Treutter, 2010). In proanthocyanidin-rich plants like tea and grapes, *F3'5'H* gene was expanded or was expressed with relatively high level (Falginella et al., 2010; Li et al., 2014; Jin et al., 2017), which is highly correlated with proanthocyanidin accumulation. ANR is a key enzyme for the biosynthesis of epicatechins in plant, which is highly correlated with proanthocyanidins accumulation as a marker gene as reported in *M. truncatula* (Pang et al., 2007; Liu et al., 2014). Therefore, it is reasonable to deduce that differentially expressed genes could be potential candidate genes for proanthocyanidin accumulation in leaves of sainfoin.

MYB transcription factors are involved in many secondary metabolic pathways (Allan and Espley, 2018; Deng et al., 2020; Chen et al., 2021), and MYB in the MBW complex is essential for the biosynthesis of proanthocyanidins in many plant species (Li, 2014; Liu et al., 2014; An et al., 2018; Wang et al., 2019). In this study, we identified one of the differentially expressed MYB genes, *OvMYBPA2*, and found that it was expressed higher in leaves of HPAs sainfoin than in LPAs sainfoin (**Figure 4**), showing high sequence similarity with other MYBs involved in proanthocyanidin pathway (**Figure 5**), such as AtTT2 from *Arabidopsis* and LjTT2a from *L. japonicus*, these genes are the key players in the regulation of proanthocyanidins (Nesi et al., 2001; Yoshida et al., 2008), and many *MYB* genes in proanthocyanidin pathway were also applied in bioengineering of proanthocyanidins (Pang et al., 2008; Hancock et al., 2012; Gourlay et al., 2020). In the bioengineering of secondary metabolism, hairy roots are an ideal system for the characterization of many functional genes (Pang et al., 2008; Verdier et al., 2012; Hao et al., 2020). In previous studies, over-expression of *AtTT2*, or *MtPAR1*, promoted the accumulation of proanthocyanidins in hairy roots of the model legume plant *M. truncatula* (Pang et al., 2008; Verdier et al., 2012; Liu et al., 2014). Since no stable transformation is available for sainfoin, therefore, we over-expressed *OvMYBPA2* in hairy roots of alfalfa, a main target plants for proanthocyanidin bioengineering. When *OvMYBPA2* was over-expressed in hairy roots of alfalfa, anthocyanins were decreased with the increase of proanthocyanidins (**Figure 6**), indicating that anthocyanins were re-directed to the proanthocyanidin branch by the over-expression of *OvMYBPA2*. Besides anthocyanins, total flavonoids were also highly increased in the *OvMYBPA2* over-expression lines (**Figure 6**), indicating *OvMYBPA2* activated entire flavonoid to provide more flux for proanthocyanidin branch. Whether or not *OvMYBPA2* could activate proanthocyanidins in alfalfa plants besides hairy roots is currently under further investigation.

With the development of biology technology, RNA-seq has been proven to be a mature technique to mine genes related to biological functions. In this study, comparative transcriptome of two sainfoin resources revealed a number of genes related to proanthocyanidin biosynthesis in leaves, which is a golden mine for further exploration of functional genes in proanthocyanidins pathway. Besides *OvMYBPA2*, other fifty-nine *MYB* genes were also differentially expressed in two sainfoin of HPAs or LPAs, which might be the key activators or suppressors for PA biosynthesis in leaves, and the regulatory mechanism of them will be investigated further.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are publicly available. This data can be found here: NCBI, OM929200 and PRJNA810561.

AUTHOR CONTRIBUTIONS

ZJ and YP designed this experiment. ZJ performed the experiments and drafted the manuscript. ZJ, YL, HH, and DY analyzed experimental data. YP revised the manuscript and supervised the study. All authors have read and agreed to the published version of the manuscript.

FUNDING

This project was supported by the Key Projects in Science and Technology of Inner Mongolia (2021ZD0031), the National Nature Science Foundation of China (31901386), and the Agricultural Science and Technology Innovation Program (ASTIP-IAS10).

ACKNOWLEDGMENTS

We sincerely thank Chunru Li for their help and suggestions on the generation of hairy roots and for Xiao Wang for her advice during the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.941918/full#supplementary-material

REFERENCES

Allan, A. C., and Espley, R. V. (2018). MYBs drive novel consumer traits in fruits and vegetables. *Trends Plant Sci.* 23, 693–705. doi: 10.1016/j.tplants.2018.06.001

Amirahmadi, A., Kazempour-Osaloo, S., Kaveh, A., Maassoumi, A. A., and Naderi, R. (2016). The phylogeny and new classification of the genus *Onobrychis* (Fabaceae-Hedysareae): evidence from molecular data. *Plant Syst. Evol.* 302, 1445–1456. doi: 10.1007/s00606-016-1343-1

An, J. P., Li, R., Qu, F. J., You, C. X., Wang, X. F., and Hao, Y. J. (2018). R2R3-MYB transcription factor MdMYB23 is involved in the cold tolerance and proanthocyanidin accumulation in apple. *Plant J.* 296: 562–577. doi: 10.1111/tpj.14050

Bai, C. K., Xu, J., Cao, B., Li, X., and Li, G. S. (2018). Transcriptomic analysis and dynamic expression of genes reveal flavonoid synthesis in *Scutellaria viscidula*. *Acta Physiol. Plant* 40, 161. doi: 10.1007/s11738-018-2733-5

Baudry, A., Heim, M. A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of *BANYULS* and proanthocyanidin biosynthesis in *Arabidopsis thaliana*. *Plant J.* 39, 366–380. doi: 10.1111/j.1365-313X.2004.02138.x

Bhattarai, S., Coulman, B., and Biligetu, B. (2016). Sainfoin (*Onobrychis viciifolia Scop.*): renewed interest as a forage legume for western Canada. *Can. J. Plant Sci.* 96, 748–756. doi: 10.1139/cjps-2015-0378

Bogs, J., Jaffe, F. W., Takos, A. M., Walker, A. R., and Robinson, S. P. (2007). The grapevine transcription factor VvMYBPA1 regulates proanthocyanidin synthesis during fruit development. *Plant Physiol.* 143, 1347–1361. doi: 10.1104/pp.106.093203

Carbonero, C. H., Mueller-Harvey, I., Brown, T. A., and Smith, L. (2011). Sainfoin (*Onobrychis viciifolia*): a beneficial forage legume. *Plant Genet. Resour.* 9, 70–85. doi: 10.1017/S1479262110000328

Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020a). TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13, 1194–1202. doi: 10.1016/j.molp.2020.06.009

Chen, J., Thilakarathna, W. P. D. W., Astatkie, T., and Rupasinghe, H. P. V. (2020b). Optimization of catechin and proanthocyanidin recovery from grape seeds using microwave-assisted extraction. *Biomolecules* 10, 243. doi: 10.3390/biom10020243

Chen, W. X., Zheng, Q. Y., Li, J. W., Liu, Y., Xu, L. Q., Zhang, Q. L., et al. (2021). DkMYB14 is a bifunctional transcription factor that regulates the accumulation of proanthocyanidin in persimmon fruit. *Plant J.* 106, 1708–1727. doi: 10.1111/tpj.15266

Cheynier, V., Comte, G., Davies, K. M., Lattanzio, V., and Martens, S. (2013). Plant phenolics: recent advances on their biosynthesis, genetics, and ecophysiology. *Plant Physiol. Biochem.* 72, 1–20. doi: 10.1016/j.plaphy.2013.05.009

Chezem, W. R., and Clay, N. K. (2016). Regulation of plant secondary metabolism and associated specialized cell development by MYBs and bHLHs. *Phytochemistry*. 131, 26–43. doi: 10.1016/j.phytochem.2016.08.006

Deng, C. P., Wang, Y., Huang, F. F., Lu, S. J., Zhao, L. M., Ma, X. Y., et al. (2020). SmMYB2 promotes salvianolic acid biosynthesis in the medicinal herb *Salvia miltiorrhiza*. *J Integr. Plant Biol.* 62, 1688–1702. doi: 10.1111/jipb.12943

Drabkova, L. Z. (2014). DNA extraction from herbarium specimens. *Methods Mol. Biol.* 1115, 69–84. doi: 10.1007/978-1-62703-767-9-4

Falginella, L., Castellarin, S. D., Testolin, R., Gambetta, G. A., Morgante, M., and Gaspero, G. D. (2010). Expansion and subfunctionalisation of flavonoid 3',5'-hydroxylases in the grapevine lineage. *BMC Genom.* 11, 562. doi: 10.1186/1471-2164-11-562

Farhadi, F., Khameneh, B., Iranshahi, M., and Iranshahy, M. (2019). Antibacterial activity of flavonoids and their structure-activity relationship: an update review. *Phytother Res.* 33, 13–40. doi: 10.1002/ptr.6208

Fujino, N., Tenma, N., Waki, T., Ito, K., Komatsuzaki, Y., Sugiyama, K., et al. (2018). Physical interactions among flavonoid enzymes in snapdragon and torenia reveal the diversity in the flavonoid metabolon organization of different plant species. *Plant J.* 94, 372–392. doi: 10.1111/tpj.13864

Gourlay, G., Ma, D. W., Schmidt, A., and Constabel, C. P. (2020). MYB134-RNAi poplar plants show reduced tannin synthesis in leaves but not roots, and increased susceptibility to oxidative stress. *J. Exp. Bot.* 71, 6601–6611. doi: 10.1093/jxb/eraa371

Hancock, K. R., Collette, V., Fraser, K., Greig, M., Xue, H., Richardson, K., et al. (2012). Expression of the R2R3-MYB transcription factor TaMYB14 from *Trifolium arvense* activates proanthocyanidin biosynthesis in the legumes *Trifolium repens* and *Medicago sativa*. *Plant Physiol.* 159, 1204–1220. doi: 10.1104/pp.112.195420

Hao, X. L., Pu, Z. Q., Cao, G., You, D. W., Zhou, Y., Deng, C. P., et al. (2020). Tanshinone and salvianolic acid biosynthesis are regulated by SmMYB98 in *Salvia miltiorrhiza* hairy roots. *J. Adv. Res.* 23, 1–12. doi: 10.1016/j.jare.2020.01.012

Hichri, I., Barrieu, F., Bogs, J., Kappel, C., Delrot, S., and Lauvergeat, V. (2011). Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J. Exp. Bot.* 62, 2465–2483. doi: 10.1093/jxb/erq442

Jin, J. Q., Ma, J. Q., Yao, M. Z., Ma, C. L., and Chen, L. (2017). Functional natural allelic variants of flavonoid 3',5'-hydroxylase gene governing catechin traits in tea plant and its relatives. *Planta* 245, 523–538. doi: 10.1007/s00425-016-2620-5

Kawai, M., Hirano, T., Higa, S., Arimitsu, J., Maruta, M., Kuwahara, Y., et al. (2007). Flavonoids and related compounds as anti-allergic substances. *Allergol. Int.* 56, 113–123. doi: 10.2332/allergolint.r-06-135

Koenen, E. J. M., Ojeda, D. I., Steeves, R., Migliore, J., Bakker, F. T., Wieringa, J. J., et al. (2020). Large-scale genomic sequence data resolve the deepest divergences in the legume phylogeny and support a near-simultaneous evolutionary origin of all six subfamilies. *New Phytol.* 225, 1355–1369. doi: 10.1111/nph.16290

Kopustinskiene, D. M., Jakstas, V., Savickas, A., and Bernatoniene, J. (2020). Flavonoids as anticancer agents. *Nutrients* 12, 457.doi: 10.3390/nu12020457

Lalani, S., and Poh, C. L. (2020). Flavonoids as antiviral agents for enterovirus A71 (EV-A71). *Viruses* 12,184. doi: 10.3390/v12020184

Lepiniec, L., Debeaujon, I., Routaboul, J. M., Baudry, A., Pourcel, L., Nesi, N., et al. (2006). Genetics and biochemistry of seed flavonoids. *Annu. Rev. Plant. Biol.* 57, 405–430. doi: 10.1146/annurev.arplant.57.032905.105252

Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* 12,323. doi: 10.1186/1471-2105-12-323

Li, Q., He, F., Zhu, B. Q., Liu, B., Sun, R. Z., Duan, C. Q., et al. (2014). Comparison of distinct transcriptional expression patterns of flavonoid biosynthesis in Cabernet Sauvignon grapes from east and west China. *Plant Physiol. Biochem.* 84, 45–56. doi: 10.1016/j.plaphy.2014.08.026

Li, S. T. (2014). Transcriptional control of flavonoid biosynthesis: fine-tuning of the MYB-bHLH-WD40 (MBW) complex. *Plant Signal. Behav.* 9, e27522. doi: 10.4161/psb.27522

Liu, C. G., Jun, J. H., and Dixon, R. A. (2014). MYB5 and MYB14 play pivotal roles in seed coat polymer biosynthesis in *Medicago truncatula*. *Plant Physiol.* 165, 1424–1439. doi: 10.1104/pp.114.241877

Liu, J. Y., Osbourn, A., and Ma, P. D. (2015). MYB transcription factors as regulators of phenylpropanoid metabolism in plants. *Mol Plant* 8, 689–708. doi: 10.1016/j.molp.2015.03.012

Liu, X. L., Hao, Y. Q., Jin, L., Xu, Z. J., Mcallister, T. A., and Wang, Y. X. (2013). Anti-Escherichia coli O157:H7 properties of purple prairie clover and sainfoin condensed tannins. *Molecules* 18, 2183–2199. doi: 10.3390/molecules18022183

Mai, L. H., Chabot, G. G., Grellier, P., Quentin, L., Dumontet, V., Poulain, C., et al. (2015). Antivascular and anti-parasite activities of natural and hemisynthetic flavonoids from new Caledonian gardenia species (Rubiaceae). *Eur. J. Med. Chem.* 93, 93–100. doi: 10.1016/j.ejmech.2015.01.012

Min, B. R., Barry, T. N., Attwood, G. T., and Mcnabb, W. C. (2003). The effect of condensed tannins on the nutrition and health of ruminants fed fresh temperate forages: a review. *Anim. Feed Sci. Technol.* 106, 3–19. doi: 10.1016/S0377-8401(03)00041-5

Mora-Ortiz, M., and Smith, L. M. J. (2018). *Onobrychis viciifolia*: a comprehensive literature review of its history, etymology, taxonomy, genetics, agronomy and botany. *Plant Genet Res.* 16, 403–418. doi: 10.1017/S1479262118000230

Nesi, N., Jond, C., Debeaujon, I., Caboche, M., and Lepiniec, L. (2001). The Arabidopsis *TT2* gene encodes an R2R3 MYB domain protein that acts as a key determinant for proanthocyanidin accumulation in developing seed. *Plant Cell*. 3, 2099–2114. doi: 10.1105/tpc.010098

Panche, A. N., Diwan, A. D., and Chandra, S. R. (2016). Flavonoids: an overview. *J. Nutr. Sci.* 5, e47. doi: 10.1017/jns.2016.41

Pang, Y. Z., Peel, G. J., Sharma, S. B., Tang, Y. H., and Dixon, R. A. (2008). A transcript profiling approach reveals an epicatechin-specific glucosyltransferase expressed in the seed coat of *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 105, 14210–14215. doi: 10.1073/pnas.0805954105

Pang, Y. Z., Peel, G. J., Wright, E., Wang, Z. Y., and Dixon, R. A. (2007). Early steps in proanthocyanidin biosynthesis in the model legume *Medicago truncatula*. *Plant Physiol.* 145, 601–615. doi: 10.1104/pp.107.107326

Pang, Y. Z., Wenger, J. P., Saathoff, K., Peel, G. J., and Wen, J. Q., Huhman,D, et al. (2009). A WD40 repeat protein from *Medicago truncatula* is necessary for tissue-specific anthocyanin and proanthocyanidin biosynthesis but not for trichome development. *Plant Physiol.* 151, 1114–1129. doi: 10.1104/pp.109.144022

Pietta, P. G. (2000). Flavonoids as antioxidants. *J. Nat. Prod.* 63, 1035–1042. doi: 10.1021/np9904509

Re, G. A., Piluzza, G., Sulas, L., Franca, A., Porqueddu, C., Sanna, F., et al. (2014). Condensed tannin accumulation and nitrogen fixation potential of *Onobrychis viciifolia Scop. grown in a Mediterranean environment. J. Sci. Food Agric.* 94, 639–645. doi: 10.1002/jsfa.6463

Regos, I., and Treutter, D. (2010). Optimization of a high-performance liquid chromatography method for the analysis of complex polyphenol mixtures and application for sainfoin extracts (*Onobrychis viciifolia*). *J. Chromatogr. A* 1217, 6169–6177. doi: 10.1016/j.chroma.2010.07.075

Regos, I., Urbanella, A., and Treutter, D. (2009). Identification and quantification of phenolic compounds from the forage legume sainfoin (*Onobrychis viciifolia*). *J. Sci. Food Agric.* 57, 5843–5852. doi: 10.1021/jf900625r

Rivaroli, D., Prunier, A., Meteau, K., Do Prado, I. N., and Prache, S. (2019). Tannin-rich sainfoin pellet supplementation reduces fat volatile indoles content and delays digestive parasitism in lambs grazing alfalfa. *Animal* 13, 1883–1890. doi: 10.1017/S1751731118003543

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., et al. (2013). The flavonoid biosynthetic pathway in *Arabidopsis*: structural and genetic diversity. *Plant Physiol. Biochem.* 72, 21–34. doi: 10.1016/j.plaphy.2013.02.001

Serafini, M., Peluso, I., and Raguzzini, A. (2010). Flavonoids as anti-inflammatory agents. *Proc Nutr Soc* 69, 273–278. doi: 10.1017/S002966511000162x

Shen, G. A., Wu, R. R., Xia, Y. Y., and Pang, Y. Z. (2021). Identification of transcription factor genes and functional characterization of *PlMYB1* from *Pueraria lobata*. *Front. Plant Sci.* 12, 743518. doi: 10.3389/fpls.2021.743518

Su, X. J., Xia, Y. Y., Jiang, W. B., Shen, G. A., and Pang, Y. Y. (2020). GbMYBR1 from *Ginkgo biloba* represses phenylpropanoid biosynthesis and trichome development in *Arabidopsis*. *Planta* 252, 68. doi: 10.1007/s00425-020-03476-1

Sun, L. W., Yu, D., Wu, Z. C., Wang, C., Yu, L., Wei, A. Z., et al. (2019). Comparative transcriptome analysis and expression of genes reveal the biosynthesis and accumulation patterns of key flavonoids in different varieties of *Zanthoxylum bungeanum* leaves. *J. Agric. Food Chem.* 67, 13258–13268. doi: 10.1021/acs.jafc.9b05732

Tohge, T., De Souza, L. P., and Fernie, A. R. (2017). Current understanding of the pathways of flavonoid biosynthesis in model and crop plants. *J. Exp. Bot.* 68, 4013–4028. doi: 10.1093/jxb/erx177

Verdier, J., Zhao, J., Torres-Jerez, I., Ge, S. J., Liu, C. G., He, X. Z., et al. (2012). MtPAR MYB transcription factor acts as an on switch for proanthocyanidin biosynthesis in *Medicago truncatula*. *Proc. Natl. Acad. Sci. U.S.A.* 109, 1766–1771. doi: 10.1073/pnas.1120916109

Waghorn, G. (2008). Beneficial and detrimental effects of dietary condensed tannins for sustainable sheep and goat production-progress and challenges. *Anim. Feed Sci. Technol.* 147, 116–139. doi: 10.1016/j.anifeedsci.2007.09.013

Wang, L. J., Lu, W. X., Ran, L. Y., Dou, L. W., Yao, S., Hu, J., et al. (2019). R2R3-MYB transcription factor MYB6 promotes anthocyanin and proanthocyanidin biosynthesis but inhibits secondary cell wall formation in *Populus tomentosa*. *Plant J.* 99, 733–751. doi: 10.1111/tpj.14364

Wu, H. F., Guo, J., Chen, S. L., Liu, X., Zhou, Y., Zhang, X. P., et al. (2013). Recent developments in qualitative and quantitative analysis of phytochemical constituents and their metabolites using liquid chromatography-mass spectrometry. *J. Pharm. Biomed. Anal.* 72, 267–291. doi: 10.1016/j.jpba.2012.09.004

Xie, D. Y., and Dixon, R. A. (2005). Proanthocyanidin biosynthesis–still more questions than answers? *Phytochemistry* 66, 2127–2144. doi: 10.1016/j.phytochem.2005.01.008

Yoshida, K., Iwasaka, R., Kaneko, T., Sato, S., Tabata, S., and Sakuta, M. (2008). Functional differentiation of *Lotus japonicus* TT2s, R2R3-MYB transcription factors comprising a multigene family. *Plant Cell Physiol.* 49, 157–169. doi: 10.1093/pcp/pcn009

Yoshida, K., Kume, N., Nakaya, Y., Yamagami, A., Nakano, T., and Sakuta, M. (2010). Comparative analysis of the triplicate proanthocyanidin regulators in *Lotus japonicus*. *Plant Cell Physiol.* 51, 912–922. doi: 10.1093/pcp/pcq067

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

ORIGINAL RESEARCH
published: 04 July 2022
doi: 10.3389/fpls.2022.955628

Molecular Cloning and Functional Characterization of a β-Glucosidase Gene to Produce Platycodin D in *Platycodon grandiflorus*

*Xinglong Su[1,2†], Fei Meng[1†], Yingying Liu[3], Weimin Jiang[4], Zhaojian Wang[1], Liping Wu[1], Xiaohu Guo[1], Xiaoyan Yao[1], Jing Wu[1], Zongping Sun[5], Liangping Zha[1,2], Shuangying Gui[1,6], Daiyin Peng[1,2,7] and Shihai Xing[1,2,8]**

[1]School of Pharmacy, Anhui University of Chinese Medicine, Hefei, China, [2]Institute of Traditional Chinese Medicine Resources Protection and Development, Anhui Academy of Chinese Medicine, Hefei, China, [3]College of Humanities and International Education Exchange, Anhui University of Chinese Medicine, Hefei, China, [4]College of Life Sciences and Environment, Hengyang Normal University, Hengyang, China, [5]Engineering Technology Research Center of Anti-aging, Chinese Herbal Medicine, Fuyang Normal University, Fuyang, China, [6]Anhui Province Key Laboratory of Pharmaceutical Preparation Technology and Application, Anhui University of Chinese Medicine, Hefei, China, [7]MOE-Anhui, Joint Collaborative Innovation Center for Quality Improvement of Anhui Genuine Chinese Medicinal Materials, Hefei, China, [8]Anhui Province Key Laboratory of Research and Development of Chinese Medicine, Anhui University of Chinese Medicine, Hefei, China

OPEN ACCESS

Edited by:
Guoyin Kai,
Zhejiang Chinese Medical University, China

Reviewed by:
Dongfeng Yang,
Zhejiang Sci-Tech University, China
Jianping Xue,
Huaibei Normal University, China

***Correspondence:**
Shihai Xing
xshshihai@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Plant Systems and Synthetic Biology, a section of the journal Frontiers in Plant Science

Received: 29 May 2022
Accepted: 16 June 2022
Published: 04 July 2022

Citation:
Su X, Meng F, Liu Y, Jiang W, Wang Z, Wu L, Guo X, Yao X, Wu J, Sun Z, Zha L, Gui S, Peng D and Xing S (2022) Molecular Cloning and Functional Characterization of a β-Glucosidase Gene to Produce Platycodin D in Platycodon grandiflorus. Front. Plant Sci. 13:955628. doi: 10.3389/fpls.2022.955628

Platycodin D (PD) is a deglycosylated triterpene saponin with much higher pharmacological activity than glycosylated platycoside E (PE). Extensive studies *in vitro* showed that the transformation of platycoside E to platycodin D can be achieved using β-glucosidase extracted from several bacteria. However, whether similar enzymes in *Platycodon grandiflorus* could convert platycoside E to platycodin D, as well as the molecular mechanism underlying the deglycosylation process of platycodon E, remain unclear. Here, we identified a β-glucosidase in *P. grandiflorus* from our previous RNA-seq analysis, with a full-length cDNA of 1,488 bp encoding 495 amino acids. Bioinformatics and phylogenetic analyses showed that β-glucosidases in *P. grandiflorus* have high homology with other plant β-glucosidases. Subcellular localization showed that there is no subcellular preference for its encoding gene. β-glucosidase was successfully expressed as 6 × His-tagged fusion protein in *Escherichia coli* BL21 (DE3). Western blot analysis yielded a recombinant protein of approximately 68 kDa. *In vitro* enzymatic reactions determined that β-glucosidase was functional and could convert PE to PD. RT-qPCR analysis showed that the expression level of β-glucosidase was higher at night than during the day, with the highest expression level between 9:00 and 12:00 at night. Analysis of the promoter sequence showed many light-responsive cis-acting elements, suggesting that the light might regulate the gene. The results will contribute to the further study of the biosynthesis and metabolism regulation of triterpenoid saponins in *P. grandiflorus*.

Keywords: *Platycodon grandiflorus*, triterpenoid saponins, β-glucosidase, subcellular localization, functional characterization

INTRODUCTION

Platycodon grandiflorus (Jacq.) A. DC. is a perennial herb of the family Campanulaceae, the roots (*Platycodi radix*) of which are used as medicinal herbs in Northeast Asia for their curative effect on respiratory diseases, including bronchitis, tonsillitis, sore throat, asthma, and tuberculosis (Lee et al., 2012; Zhang et al., 2015). Oleanane-type triterpenoid saponins, especially platycodin D [formal name: 3β-(β-D-glucopyranosyloxy)-2β, 16α, 23, 24-tetrahydroxy-O-D-apio-β-D-furanosyl-(1→3)-O-β-D-xylopyranosyl-(1→4)-O-6-deoxy-olean-12-en-28-oic acid, PD], are the major active components in *P. grandiflorus*. PD exhibits a wide range of pharmacological effects, including anti-atherosclerosis (Wu et al., 2012), anti-obesity (Lee et al., 2012), anti-inflammation (Guo et al., 2021), anti-oxidant (Wang et al., 2018), anti-aging (Shi et al., 2020), and anti-cancer (Huang et al., 2019) effects, as well as possesses excellent application and development potential.

PD consists of a main oleanane-type backbone with C3-Glc and C28-O-Api-Xyl-Rha-Ara. In addition, platycoside E (PE) is a major platycoside, accounting for more than 20% content of the *platycodi radix* platycosides (Yoo et al., 2011). The chemical structure of PE includes only two more glucosyl groups (at the C3 position) than PD (**Supplementary Figure 1**). Leaves of *P. grandiflorus* are rich in PE, which were discarded when collecting medicinal plants. (Shin et al., 2019; Su et al., 2021; Yu et al., 2021). Reducing the PE content-a lot of which is wasted a lot in actual production-may increase the PD content of *P. grandiflorus* plants. The pharmacological activities of deglycosylated ginsenosides are considerably higher than those of glycosylated ginsenosides (Shin and Oh, 2016), owing to theirs' lower molecular weight, better hydrophobicity, and easier absorption in the human gastrointestinal tract. Therefore, several studies have investigated the deglycosylation of platycosides using various methods. Biotransformation, especially, displays the highest selectivity and productivity. PD is a deglycosylated triterpene saponin, whereas PE is a glycosylated triterpene saponin.

β-glucosidases (β-D-glucoside glucohydrolase, EC 3.2.1.21) are involved in diverse cellular functions (e.g., they hydrolyze the glycosidic bonds of alkyl-, amino-, or aryl-β-D glucosides, cyanogenic glucosides, disaccharides, and short oligosaccharides), and are ubiquitous enzymes found in archaea, eubacteria, and eukaryotes (Bhatia et al., 2002; Ketudat Cairns and Esen, 2010). These enzymes are involved in glycolipids breakdown and glucose release in plants to meet their metabolic needs (e.g., for cell growth and wall remodeling; Chuenchor et al., 2008; Horikoshi et al., 2022). β-glucosidases also display a range of aglycone specificities, a few of which show almost absolute specificity for one sugar and one aglycone, while others accept a range of either glycones or aglycones, or both (Baglioni et al., 2021; Dziadas et al., 2021; Prieto et al., 2021). They have been extensively studied in the biotransformation of PE to PD. The enzymatic biotransformation of PE to PD *in vitro* has been demonstrated in various studies using β-galactosidase from *Aspergillus oryzae* (Ha et al., 2010), snailase (Li et al., 2012b), laminarinase (Jeong et al., 2014), β-glucosidase from *Aspergillus usamii* (Ahn et al., 2018), recombinant β-glucosidases from *Caldicellulosiruptor bescii* (Kil et al., 2019), and cytolase (Shin et al., 2020). We further speculated that there might be a deglycosylase in *P. grandiflorus*, which could also catalyze the conversion of PE into PD (Su et al., 2021). Although great progress has been made in the study of how enzymes catalyzed the conversion of PE to PD, little is known about the β-glucosidase gene with this function in *P. grandiflorus*. Therefore, the present study aimed to investigate the β-glucosidase gene from *P. grandiflorus* using complementary DNA (cDNA) cloning and *in vitro* enzymatic characterization. Our findings may provide new insights into the analysis and regulation of the PD biosynthetic pathway in *P. grandiflorus*.

MATERIALS AND METHODS

Plant Materials

Platycodon grandiflorus plants used in this study were the same materials described by Su et al. (2021). The specimen of *P. grandiflorus* was deposited in the Herbarium of Anhui University of Chinese Medicine (depository number 20200705). Samples were collected at 6:00 AM, 9:00 AM, 12:00 AM, 3:00 PM, 6:00 PM, 9:00 PM, and 12:00 PM, 3 days in a row and immediately stored in liquid nitrogen for subsequent total RNA isolation. Three biological replicates were performed for each sample.

Total RNA Preparation, cDNA Synthesis and RT-qPCR

Total RNA was extracted using TRNzol Universal Reagent (Tiangen Biotech Co., Ltd., Beijing, China) according to the manufacturer's instructions. Each total RNA sample was qualified and quantified using the ultramicro spectrophotometer DS-11 (DeNovix, Wilmington, Delaware, USA). cDNA was synthesized using the FastKing RT Kit (Tiangen Biotech Co., Ltd., Beijing, China). SuperReal PreMix Plus (Tiangen Biotech Co., Ltd., Beijing, China) was used for RT-qPCR on the LightCycle480 platform (Roche, Switzerland) to determine the mRNA transcriptional levels of the genes encoding β-glucosidase and β-amyrin synthase (β-AS, GenBank: KY412556.1) in *P. grandiflorus*. The mRNA of *18sRNA* was used as an internal reference. Three biological replicates were used, and the relative expression of the mRNA level was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001). The sequence of the gene encoding β-glucosidase, its promoter sequence, and all the primer pairs used are listed in **Supplementary Tables 1** and **2**.

Bioinformatics Analysis

Previously, the cDNA sequence of the candidate gene encoding β-glucosidase was obtained *via* transcriptomic sequencing (Su et al., 2021). ExPASy[1] was used to deduct the amino acid (aa) sequences and predict physicochemical properties. The conserved domains of β-glucosidase were detected using Multiple Em

[1]http://www.expasy.ch/tools/

for Motif Elicitation[2] and the NCBI Conserved Domain Database (CDD)[3] search tools. The possible transmembrane regions of the protein were analyzed using the TMHMM online program.[4] PSORT[5] was used to predict the subcellular localization of the protein encoded by the candidate gene. The secondary structure and three-dimensional homologous modeling of the protein were predicted using softwares of GOR IV,[6] PDBsum,[7] and SWISS-MODEL,[8] respectively. Online PlantCARE[9] was used to analyze the cis-acting elements of promoters. The amino acid sequence was used to perform homology searched on NCBI, and the sequences with high similarity were selected for multiple alignment. Phylogenetic analysis was performed using the MEGA X software (maximum likelihood), and using bootstrap analysis with 1,000 iterations.

Subcellular Localization

The full-length cDNA sequence of the candidate gene encoding β-glucosidase without the terminator codons was inserted into the pBI121-EGFP vector to generate the pBI121-β-glucosidase-EGFP fusion protein, which, after sequencing, was subsequently transformed into *Agrobacterium tumefaciens* GV3101 using the freeze–thaw method. Two batches of *A. tumefaciens* containing pBI121-β-glucosidase-EGFP and pBI121-EGFP, respectively, was cultured overnight until the OD600 was approximately 1.0. Next, these batches of *A. tumefaciens* were resuspended in infiltration buffer (10 mM $MgCl_2$, 10 mM MES (pH 5.6), and 100 μM acetosyringone) to an OD600 of 0.6 and incubated at room temperature for 2 h before being syringe-infiltrated into 5-week-old *Nicotiana benthamiana* leaves. Signals were observed under a ZEISS710 confocal laser scanning microscope (ZEISS, Germany) 48 h after infiltration.

Preparation of Recombinant β-Glucosidase

β-glucosidase-3-F and β-glucosidase-3-R primers were designed to introduce new restriction sites by changing nucleotides (**Supplementary Table 3**). The target fragments were amplified using Ex Taq DNA polymerase (Takara, Biomedical Technology Co., Ltd., Beijing, China) *via* PCR, double-restricted using Bgl II/Xho I (NEB, Beijing, China), and ligated to the Bgl II/Xho I double-restricted pET-32α(+) vector using T4 DNA Ligase (NEB, Beijing, China). The recombinant vector was transformed into *E. coli* DH5α cells using the freeze–thaw method, and positive transformants were selected on Lysogeny Broth (LB) agar plates containing ampicillin (final concentration: 50 μg/ml) for expansion and sequencing. *E. coli* BL21 (DE3) was then transformed with the recombinant plasmid for protein expression.

[2]https://meme-suite.org/meme/tools/meme
[3]http://www.ncbi.nlm.nih.gov/cdd/
[4]http://www.cbs.dtu.dk/services/TMHMM-2.0
[5]https://www.genscript.com/psort.html
[6]https://npsa-prabi.ibcp.fr/
[7]https://www.ebi.ac.uk/pdbsum/
[8]https://swissmodel.expasy.org/
[9]http://bioinformatics.psb.ugent.be/webtools/plantcare/html/

The transformed BL21 (DE3) cells were incubated in LB medium with 100 μg/ml ampicillin at 37°C until OD_{600} reached 0.6–0.8. After adding IPTG (isopropyl-β-D-thiogalactoside) at a final concentration of 0.3 mM to the medium, the cells were further cultured on a rotary shaker at 18°C for 18 h, then harvested and re-suspended in phosphate buffer (0.01 M, pH 7.2). Phenylmethyl sulfonyl fluoride (final concentration: 0.5 mM) and lysozyme (final concentration: 0.3 mg/ml) were added to the resuspension, followed by sonication at 4°C. The supernatant was collected *via* centrifugation, and the protein concentration was determined using the Modified BCA Protein Assay Kit (Shanghai Sangon Biotech, Shanghai, China). Proteins (20 μg) were separated using 12% SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and visualized by staining with Coomassie Brilliant Blue R-250.

Western Blot Analysis

Each protein sample was resolved separated using SDS-PAGE, transferred onto polyvinylidene fluoride (PVDF) membranes (Beijing Labgic technology Co., Ltd., Beijing, China), and blocked with 5% bovine serum albumin in TBST [20 mM Tris–HCl (pH 7.4), 150 mM NaCl, and 0.1% Tween 20] for 2 h at room temperature. The membrane was washed with TBST and incubated overnight at 4°C in a 1/3,000 dilution of anti-6 × His tag mouse monoclonal antibody (BBI life sciences, China). After three washes with TBST at 10 min intervals, the membrane was incubated in a 1/5,000 dilution of AP-conjugated rabbit anti-mouse IgG (BBI Life Sciences, China) at room temperature. After three washes with TBST, two washes with TBS [20 mM Tris–HCl (pH 7.4) and 150 mM NaCl], and one wash with AP buffer [100 mM NaCl, 50 mM Tris–HCl (pH 8.0), and 2 mM $MgCl_2$], recombinant protein was visualized *via* BCIP/NBT alkaline phosphatase staining.

In vitro Enzymatic Reactions

Four experimental groups were established to verify the function of putative β-glucosidase, namely, the crude enzymes from *E. coli* BL21 (DE3), the crude enzymes from *E. coli* BL21 (DE3) containing the pET-32α(+) vector, the crude enzymes from *E. coli* BL21 (DE3) including the pET-32α(+)-β-glucosidase recombinant vector, and boiled crude enzymes from *E. coli* BL21 (DE3) with the pET-32α(+)-β-glucosidase recombinant vector. PE (final concentration: 0.4 mg/ml) and each crude enzyme extract (final concentration: 0.05 mg/ml) were mixed in a phosphate buffer (0.01 M, pH 7.2) and incubated at 37°C for 2 h. The reaction products were analyzed using the LC-16 high-performance liquid chromatography (HPLC) system (Shimadzu, Japan). A Topsil C18 column (4.6 mm × 250 mm; Agilent, USA) was used with the isocratic elution of water and acetonitrile (71:29). The flow rate was 1 ml/min with a detection wavelength of 210 nm, and the column compartment was maintained at 30°C (Su et al., 2021).

Statistical Analysis

All experiments were performed independently at least three times, and the data are expressed as the mean ± standard error.

GraphPad (version 8.0) suite was used for statistical analysis, and $p \leq 0.05$ was deemed a statistically significant difference.

RESULTS

Cloning and Analysis of the β-Glucosidase Encoding Gene

We previously obtained three candidate genes (CL4020.Contig1_All, Unigene 1627_All, and Unigene7900_All; Su et al., 2021). In the present study, we selected CL4020.Contig1_All with the highest similarity as the candidate gene of *Pgβ-glucosidase* for our investigation. The full length of the candidate β-glucosidase gene was amplified using the β-glucosidase-F/β-glucosidase-R primers (**Supplementary Table 1**), and the gene sequence was submitted to GenBank (OM867671) in NCBI. The resulting PCR products were subjected to 1% agarose gel electrophoresis, and the results are shown in **Supplementary Figure 2**. The full-length cDNA putatively encoding β-glucosidase is 1,488-base pair (bp)-long and encodes a polypeptide with 495 amino acids (aa). In addition, the molecular mass of β-glucosidase was found to be 56,479.22 Da, the theoretical isoelectric points were 5.16, and the half-life was estimated to be 30 h (mammalian reticulocytes, *in vitro*). The instability index (II) was computed to be 30.97 and was classified as stable. Conserved domain analysis revealed that the protein belongs to the glycosyl hydrolase family 1 (GH1), and the actual alignment was detected with superfamily member pfam00232 (**Supplementary Figure 3**). The predicted protein secondary domain shows that the protein amino acids mainly exist as alpha helices, extended strands, and random coils. Furthermore, according to the template 3gno.1.A model, the global model quality estimation (GMQE) value is 0.82 (**Supplementary Figure 4**). From the predicted results, the protein has no transmembrane regions.

Phylogenetic Tree and Conserved Domain Analysis

The phylogenetic tree of 12 different plants and two bacterial species was constructed, including *P. grandiflorus* (OM867671), *Nicotiana attenuata* (XP_019261514.1), *Nicotiana sylvestris* (XP_009767051.1), *Solanum tuberosum* (XP_006353254.1), *Handroanthus impetiginosus* (PIN23194.1), *Actinidia chinensis* var. *chinensis* (PSR98148.1), *Camellia sinensis* (XP_028088410.1), *Helianthus annuus* (XP_022015860.1), *Erigeron canadensis* (XP_043617153.1), *Cynara cardunculus* var. *scolymus* (XP_024969531.1), *Lactuca sativa* (XP_023733473.1), *Pyrus x bretschneideri* (XP_018501441.1), *C. bescii* DSM 6725 (ACM59590), and *Caldicellulosiruptor owensensis* OL (ADQ03897). Notably, β-glycosidases from *C. bescii* and *C. owensensis* have converted PE to PD (Kil et al., 2019; Shin et al., 2019). Compared with *P. grandiflorus*, the relationship between the β-glucosidase of *Pyrus x bretschneideri* and that of the two bacterial species was close. Four β-glucosidases (XP_022015860.1, XP_043617153.1, XP_022015860.1, and XP_023733473.1) from Asteraceae family members were phylogenetically closest to that of *P. grandiflorus* (**Figure 1A**). Furthermore, from the perspective of the evolutionary relationship of the genome, *H. annuus* is very close to *P. grandiflorus* and, they both belong to Asterids II (Kim et al., 2020).

Multiple sequence alignment revealed that Pgβ-glucosidase contained two conserved carboxylic acid residues (E192 and E393), serving as the catalytic acid–base and nucleophile (Henrissat et al., 1995; Jenkins et al., 1995; Chuenchor et al., 2008; **Figure 2**). These enzymes present a catalytic cycle that occurs in two distinct steps (glycosylation and deglycosylation), and where the two active sites involved are critical for their double displacement (Zechel and Withers, 2000; Marana, 2006; Chuenchor et al., 2011). Moreover, conserved domain analysis revealed that the β-glucosidases in other plants all possess motif 1, motif 2, motif 3, motif 4, and motif 5, except for those of *C. bescii* and *C. owensensis*, which only have motif 1, motif 2, and motif 3 (**Figure 1B**). However, β-glucosidase from *C. bescii* and *C. owensensis* have been shown to be function in catalyzing the conversion of PE to PD, thus, the common motif 1, motif 2, and motif 3 may have critical roles. In addition, E238, L298, and A425 in Pgβ-glucosidase were highly consistent with the β-glucosidases derived from *C. bescii* and *C. owensensis* and were different from other species (**Figure 2**).

Subcellular Localization

Prediction of subcellular localization suggested that Pgβ-glucosidase was localized in the cytoplasm, mitochondria, and nucleus. GFP was fused to the C-terminal domain of the β-glucosidase protein and transiently transformed in *N. benthamiana* to determine the subcellular localization of β-glucosidase in *P. grandifloras*. Subsequently, confocal laser scanning microscopy was used to observe the *Agrobacterium*-infected the leaves of *N. benthamiana* containing the gene of interest, and the results are shown in **Figure 3**. Strong fluorescence signals were distributed in the cytoplasm and nucleus, consistent with the prediction. The results indicated that β-glucosidase was localized in the cytoplasm and nucleus of *P. grandiflorus*.

Expression and Functional Assay of β-Glucosidase

Inducible expression of β-glucosidase recombinant protein was performed in pET-32α(+) using Trx, a 6×His and an S fusion tag at the N-terminal. With the size increase of 16.3 kDa by the fusion tags, the expected β-glucosidase protein was approximately 72.7 kDa (**Figure 4**). In addition to the expression of the recombinant plasmid (lane 3), *E. coli* BL21 (DE3) without the plasmid (lane 1) and *E. coli* BL21 (DE3) with the pET-32α(+) vector (lane 2) were used as controls (**Figure 4**). In addition *E. coli* BL21 without the plasmid (lane 1) and *E. coli* BL21 with the pET-32α(+) vector (lane 2) were performed as controls. Following crude protein quantification, 20 μg of proteins from each group was loaded separately and subjected to SDS-PAGE analysis. A thick band of recombinant β-glucosidase protein in lane 3 was in approximately 73.0 kDa (**Figure 4A**). Subsequently, western blotting using an anti-6×His tag mouse monoclonal antibody showed a band of approximately 73.0 kDa on the PVDF

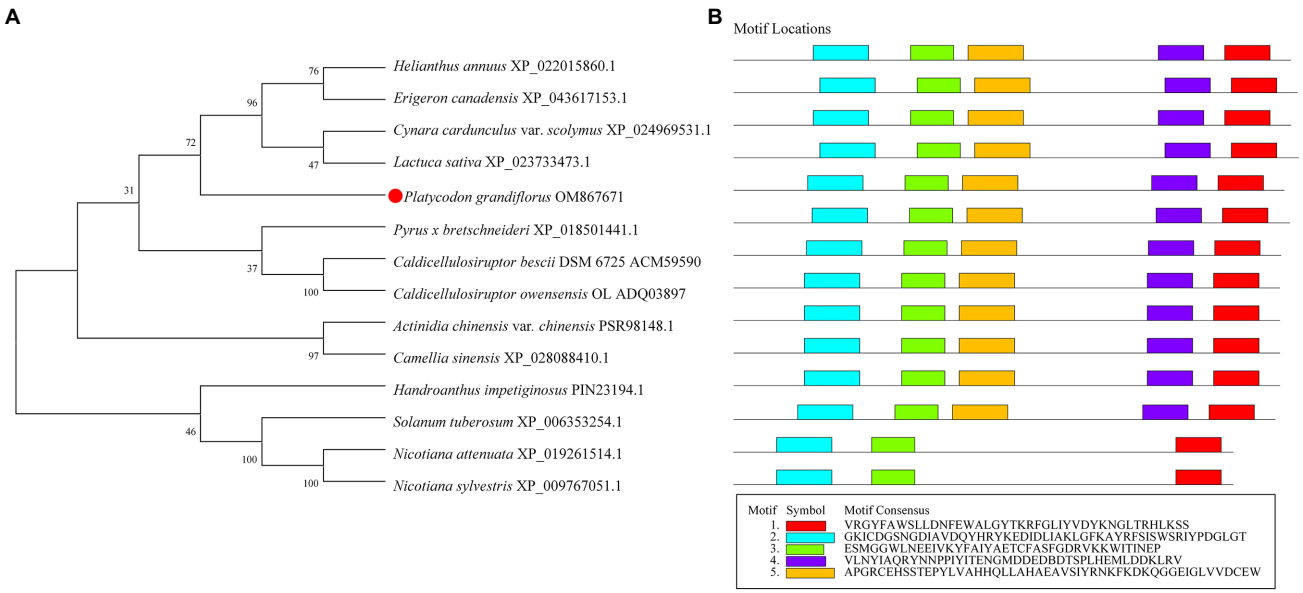


FIGURE 1 | Phylogenetic and conserved domain analysis of Pgβ-glucosidase with those of other species. **(A)** Neighbor-joining phylogenetic tree, constructed using MEGA X. **(B)** Conserved domains, detected using Multiple Em for Motif Elicitation.

membrane in lane 3, whereas lanes 1 and 2 (controls) showed no band. This finding confirmed the successful expression of β-glucosidase after transformation (**Figure 4B**).

The products of a putative conversion of PE to PD were analyzed using HPLC to verify the function of the candidate β-glucosidase. The reaction mixture, consisting of *E. coli* crude protein extracts (final concentration: 0.05 mg/ml) and PE (final concentration: 0.4 mg/ml) was incubated in a water bath at 37°C, and analyzed *via* HPLC after a 2 h reaction in 0.01 M phosphate buffer (pH 7.2). To test whether the β-glucosidase recombinant protein could catalyze the conversion of PE to PD *in vitro*, we incubated *E. coli* BL21 (DE3) protein extracts (BL21) and *E. coli* BL21 (DE3) with the pET-32α(+) vector (BL21 + pET-32α) under the same conditions. In addition, the experimental group (BL21-pET-32α-β-glucosidase) protein was boiled to form a new control (BL21-pET-32α-β-glucosidase boiled). Standard compounds of PD (> 98% purity) and PE (> 98% purity) were purchased from Chengdu Push Bio-Technology Co., Ltd., and the retention time and peak profile of 0.4 mg/ml of standard PD and 0.4 mg/ml of standard PE are shown in **Figure 5A**. Our results revealed that only the experimental group (BL21-pET-32α-β-glucosidase) consumed PE and generated a new substance (PD; **Figure 5**). After excluding various effects of the control groups, we believe that recombinant target candidate β-glucosidase from *P. grandiflorus* can convert PE to PD *in vitro*.

Analysis of β-Glucosidase Diurnal Expression

Studies have shown that the content of PE is higher in the leaves of *P. grandiflorus* than in other parts (Su et al., 2021; Yu et al., 2021). Therefore, we collected *P. grandiflorus* leaves at seven time points of the day and night from 6:00 AM to 12:00 PM (at 3-h intervals) as samples for analysis. In addition, we simultaneously analyzed the expression of β-amyrin (β-AS), which is considered a key branching enzyme for generating oleanane-type triterpene backbones (Kim et al., 2020; **Supplementary Figure 5**). In other words, β-amyrin is a key enzyme in the biosynthetic pathway for synthesizing PE and PD in *P. grandiflorus*. The results of RT-qPCR analysis showed that the expression of β-amyrin remained unchanged from 6:00 AM to 12:00 PM and then showed an upward trend. Furthermore, the expression was the highest at 6:00 PM and decreased after 6:00 PM; however, it was still higher than that during most times of the day (**Figure 6**). Similarly, the expression level of β-glucosidase is higher at night than during the daytime, with the highest values recorded from 9:00 to 12:00 at night. Furthermore, we obtained approximately 2000 bp of the promoter sequence (**Supplementary Table 1**) of the β-glucosidase gene from the *P. grandiflorus* genome (GenBank: GCA_016624345.1). Analysis of the promoter sequence of the β-glucosidase gene revealed the present of many light-responsive cis-acting elements such as AE-box, Box 4, GT1-motif, MRE, TCCC-motif, TCT-motif, and other cis-acting elements (**Supplementary Figure 6**). These results suggest that β-glucosidase gene expression may be regulated by light.

DISCUSSION

Platycodon grandiflorus is a common natural medicinal plant with a history of thousands of years that contains a wide variety of more than 80 chemical components, including saponins, alkynes, lipids, and flavonoids (Wang et al., 2017; Huang et al., 2021; Zhang et al., 2022). Modern pharmacological studies

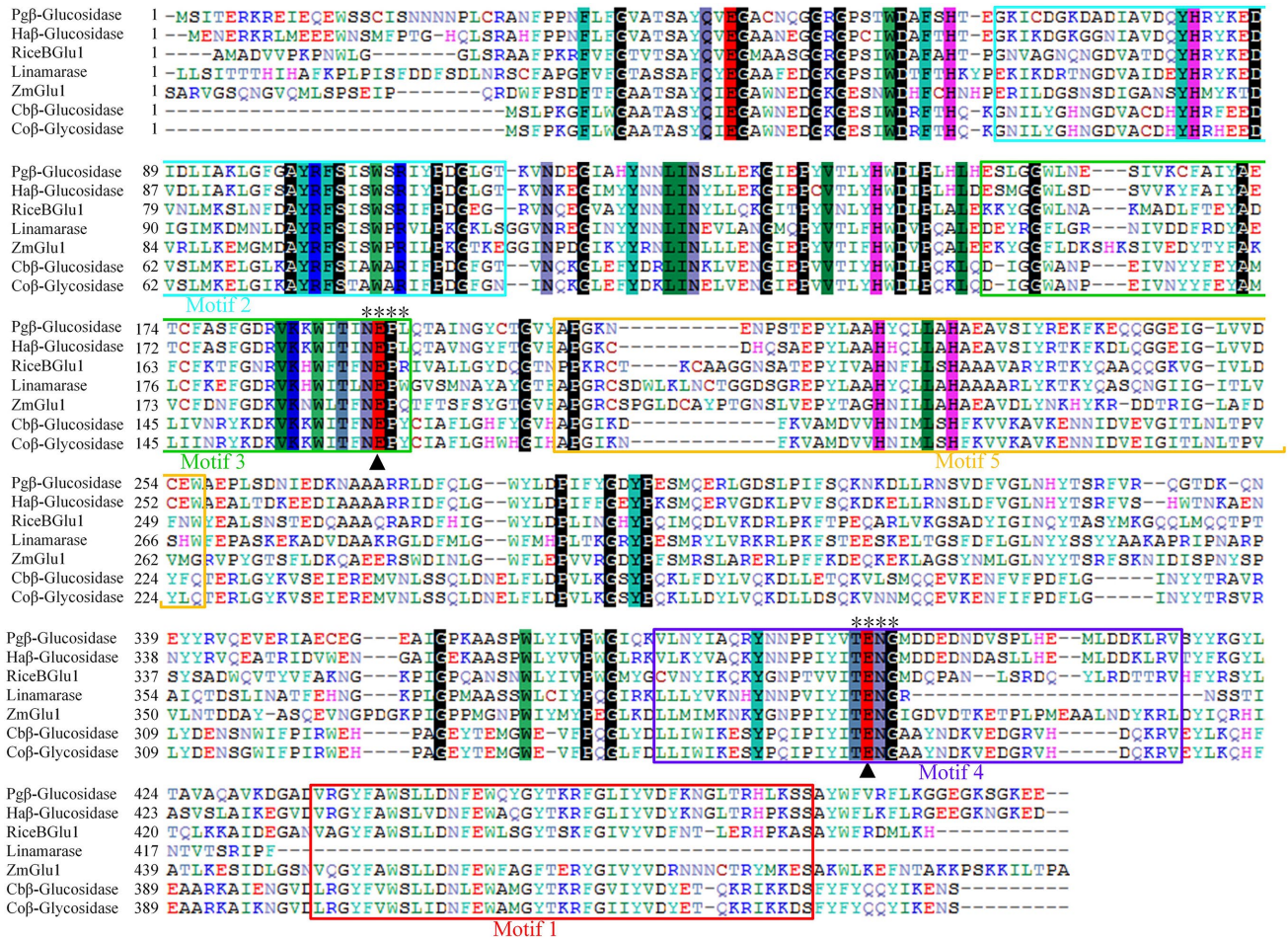


FIGURE 2 | Amino acid sequences alignment of seven β-glucosidases. Black triangles indicate acid/base and nucleophilic catalytic residues. Two motifs (TENG and NEPX) are marked out with "*." Pgβ-glucosidase, β-glucosidase of *Platycodon grandiflorus*; Haβ-glucosidase, β-glucosidase of *Helianthus annuus* (XP_022015860.1); RiceBGlu1, β-glucosidase of rice (PDB: 2RGL_A); Linamarase (GenBank: X56733.1); ZmGlu1, β-glucosidase of *Zea mays* (PDB: 1E1E_A); Cbβ-glucosidase, β-glucosidase of *Caldicellulosiruptor bescii* DSM 6725 (GenBank: ACM59590); Coβ-glucosidase, β-glucosidase of *Caldicellulosiruptor owensensis* OL (GenBank: ADQ03897).

indicated that PD is one of the main active saponins of *P. grandiflorus* with broad bioactivities (Kwon et al., 2017; Zhang et al., 2017; Cho et al., 2018). The Chinese Pharmacopoeia also used PD content as the standard to measure the qualification of medicinal materials. In the present study, we identified a glycosyltransferase β-glucosidase from *P. grandiflorus* that can remove two glucose groups from PE to generate PD with stronger activity. Furthermore, the results of the phylogenetic analysis showed that Pgβ-glucosidase is the closest to *H. annuus* in the Compositae family, which is consistent with previous findings by Kim et al. (2020) on the evolutionary relationship of the *P. grandiflorus* genome. The subcellular localization results revealed that Pgβ-glucosidase is located in the nucleus and cytoplasm of *P. grandiflorus*. The results of the 1-day expression pattern and promoter sequence analysis indicated that the transition of related secondary metabolites might mainly occur at night; therefore, it is speculated that the gene encoding Pgβ-glucosidase may be regulated by light.

Analyzing known functionally related sequences for common sequence domains or motifs can reveal their association with a common function (Stormo, 2006). For example, the amino acid sequence of chalcone synthase from *Coelogyne ovalis* Lindl. shares four motifs with chalcone synthase from the *Oncidium* hybrid cultivar., *Cymbidium* hybrid cultivar., *Bromheadia finlaysoniana*, and *Dendrobium nobile*, which contains the active site "RLMLYQQGCFA GGTVLR" and the signature sequence of "GVLFGFGPGL" of the chalcone synthase protein (Singh and Kumaria, 2020). In addition, a study of β-glucosidase found that most of its homologous sequences contained the conserved TENG and NEPW motifs, which contained two conserved glutamic acid residues as acid/base catalyst and an active catalytic nucleophile, respectively (Zada et al., 2021a,b). We found five conserved motifs that also included the conserved TENG and NEPX motifs of β-glucosidases. The results of multiple sequence alignment revealed that E192 and E393 are highly

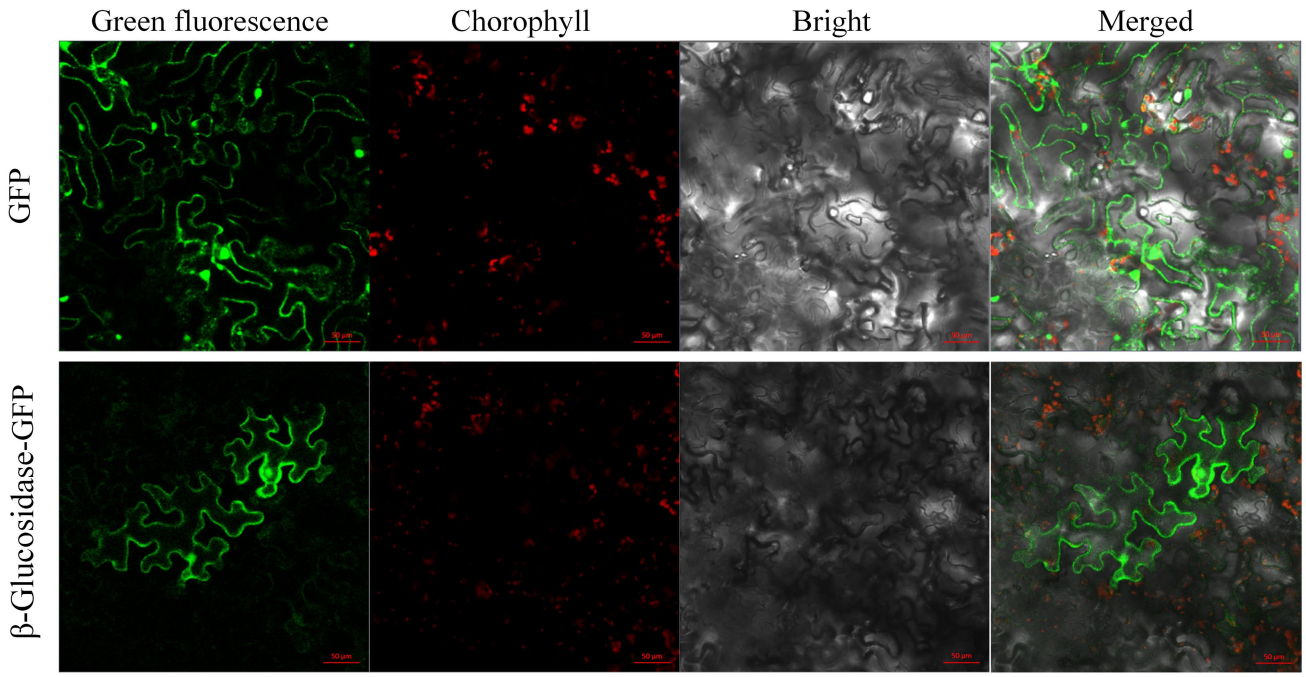


FIGURE 3 | The subcellular localization of genes encoding Pgβ-glucosidase. The subcellular localization of 35S:: GFP, and 35S:: Pgβ-glucosidase-GFP in leaf epidermal cells of *N. benthamiana* leaf epidermal cells after 48 h infiltration; the epidermal cells of *N. benthamiana* were used for taking images of green fluorescence, chloroplast autofluorescence, visible light, and merged visible light.

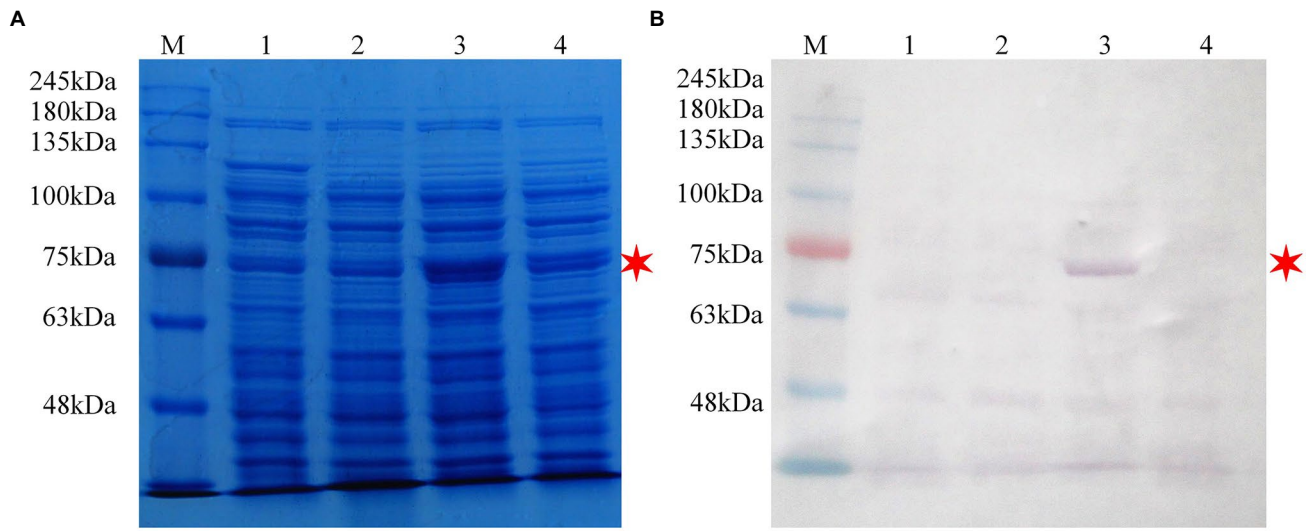


FIGURE 4 | SDS-PAGE and western blot analysis of the recombinant protein pET-32a(+)-Pgβ-glucosidase in *Escherichia coli* BL21 (DE3). **(A)** SDS-PAGE analysis. **(B)** Western blot analysis. M: protein molecular weight standards; Lane 1: cell lysate of *E. coli* BL21 (DE3; 0.3 M IPTG-inducing); Lane 2: cell lysate of *E. coli* BL21 (DE3)/pET-32a(+; 0.3 M IPTG-inducing); Lane 3: cell lysate of *E. coli* BL21 (DE3)/pET-32a(+)-Pgβ-glucosidase (0.3 M IPTG-inducing); Lane 4: cell lysate of *E. coli* BL21 (DE3)/pET-32a(+)-Pgβ-glucosidase (1.0 M IPTG-inducing). Bacterial culture conditions were kept consistent.

conserved in the β-glucosidase sequences of seven creatures, including two bacterial species (*C. bescii* and *C. owensensis*). This finding suggests that Pgβ-glucosidase may also deglycosylate PE into PD through a double displacement reaction.

β-glucosidases primarily catalyze the removal of terminal non-reducing β-D-glucosyl residues from various glucoconjugates, including glucosides, oligosaccharides, and 1-O-glucosyl esters, as well as perform transglycosylation and reverse hydrolysis. Due to the extensive distribution of their

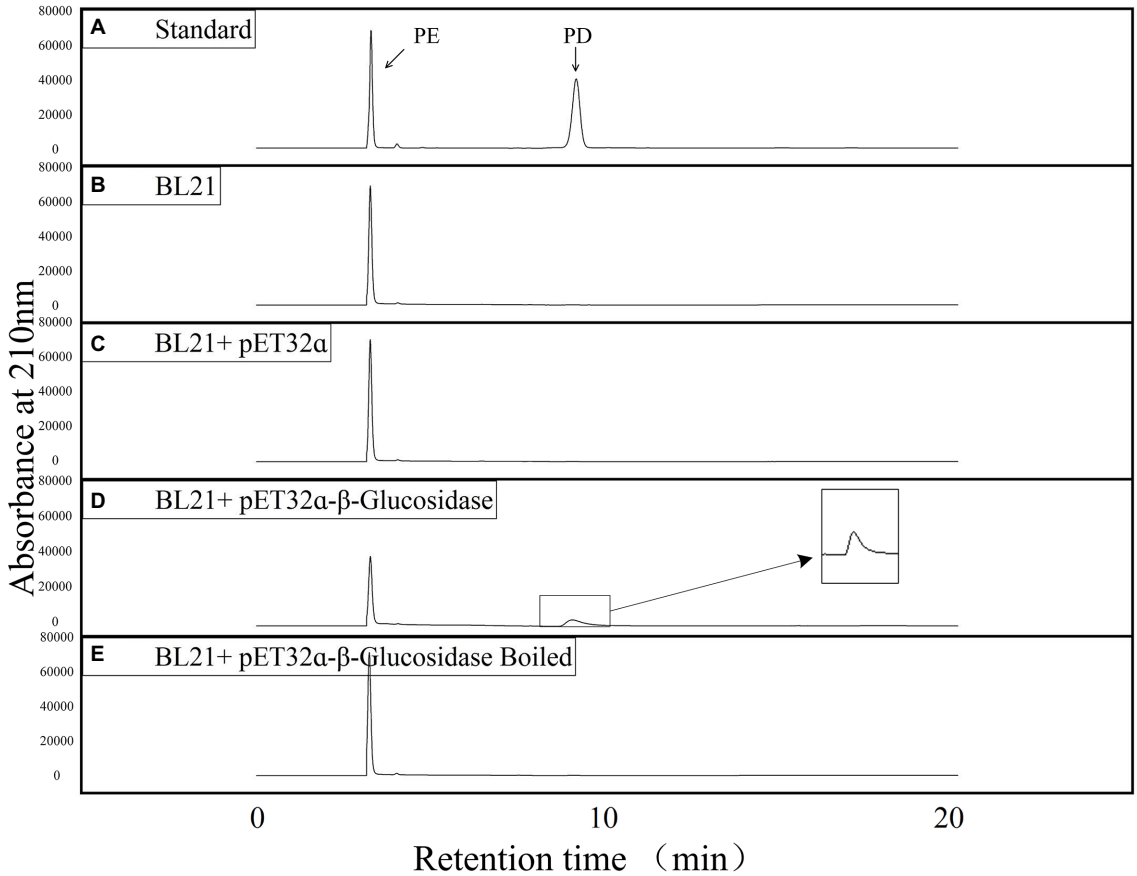


FIGURE 5 | HPLC detection of the enzyme activities of putative Pgβ-glucosidase in producing PD. **(A)** Liquid chromatograms of standard PE and PD. **(B)** Liquid chromatogram of *E. coli* BL21 (DE3) cell lysate reacting with PE. **(C)** Liquid chromatogram of *E. coli* BL21 (DE3)/pET-32α(+)cell lysate reacting with PE. **(D)** Liquid chromatogram of *E. coli* BL21 (DE3)/pET-32α(+)-Pgβ-glucosidase cell lysate reacting with PE. **(E)** Liquid chromatogram of boiled *E. coli* BL21 (DE3)/pET-32α(+)-Pgβ-glucosidase cell lysate reacting with PE.

substrates, these enzymes exist widely in nature and exhibit a range of functions (Godse et al., 2021). Studies have shown that β-glucosidase promotes the formation of free terpenes, phenylpropenes, and specific aliphatic esters during wine fermentation and promotes the production of wine aroma compounds that affect the aroma and flavor of the product (Liu et al., 2021). β-glucosidase was identified from mangrove soil showed high hydrolyzing ability for soybean isoflavone glycosides *via* heterologous expression in *E. coli*, which could completely convert daidzin and genistin to daidzein and genistein, respectively (Li et al., 2012a). Studies have also shown that β-glucosidase from the thermophilic fungus *Talaromyces leycettanus* JM12802 could hydrolyze isoflavone glycosides to aglycones (Li et al., 2018). Furthermore, recombinantly expressed β-glucosidase from *Sulfolobus solfataricus* and *Microbacterium esteraromaticum* transformed ginsenoside Rb1 to the more stable and readily absorbed, as well as more pharmacologically active, ginsenoside compound K and ginsenoside 20(S)-Rg3, respectively (Noh et al., 2009; Li et al., 2012a). β-glucosidase may also be an economically viable option for industrial use, with the production of pharmaceutically important compounds. Our current study identified a β-glucosidases from the *P. grandiflorus* plant that can convert PE into the more stable, smaller molecular weight, and more pharmacologically active PD, providing new insights for the analyzing the biosynthetic pathway of triterpenoid saponins in *P. grandiflorus*. Our findings will also lay a foundation to increase PD accumulation through molecular biological approaches, improve the quality of medicines, and expand the resources of PD.

Light is an important environmental factor for plant growth and development. It provides essential light energy for plant growth and acts as an environmental signal to regulate plant development. The downstream regulatory factors in the light signal transduction pathway mainly interact with specific cis-acting elements in the promoters of light-controlled genes, thereby up-regulating or down-regulating the expression of specific genes. The cis-acting elements present in the promoters of light-controlled genes are called light-responsive cis-elements (LREs; Hiratsuka and Chua, 1997). A previously study has shown that *SmCP*, the gene encoding *Solanum melongena* cysteine proteinase, showed maximum expression in the dark;

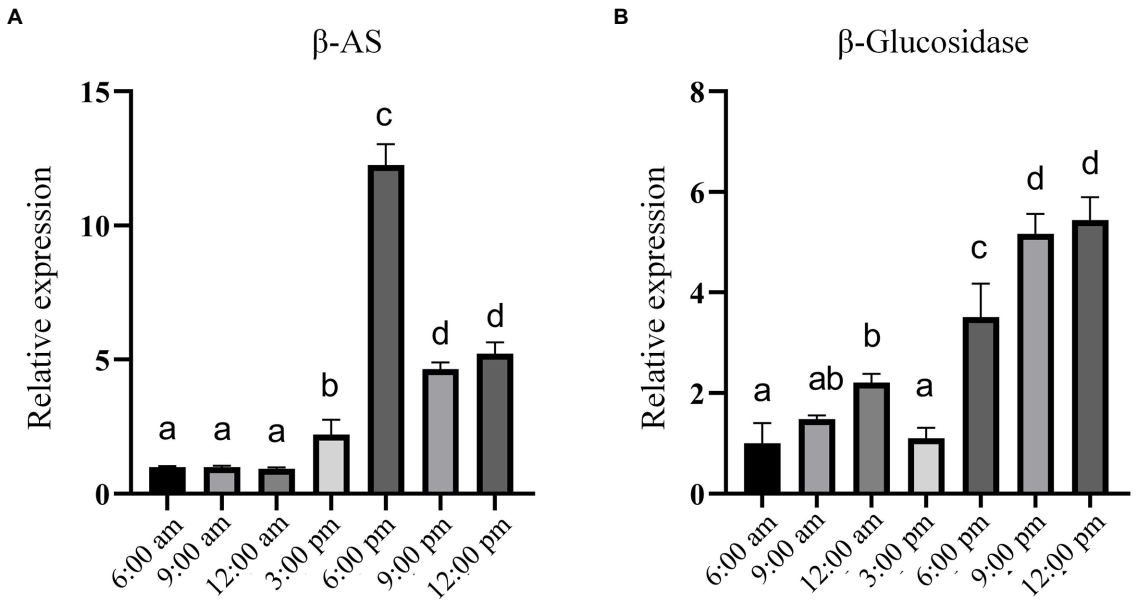


FIGURE 6 | RT-qPCR analysis of the relative expression of genes encoding β-AS and Pgβ-glucosidase at different time points within 1 day. **(A)** Relative expression of *β-AS* gene. **(B)** Relative expression of the gene encoding Pgβ-glucosidase. Error bars indicate SD ($n = 3$) and different letters represent a value of $p \leq 0.05$.

the correlation between binding activity and expression suggests that the regulation of *SmCP* is accomplished by binding to the G-box in its promoter region (Xu et al., 2003). Different flanking sequences around the light-responsive G-box core can mediate induction or inhibition (Hudson and Quail, 2003). The study of a GT element showed that it can activate the expression of the rice *phyA* gene in the dark (Dehesh et al., 1990) and that it has two functions of positive and negative regulation (Mehrotra and Panwar, 2009). The AT-rich sequence of *oat* as a negative regulatory element can reduce the expression of light-regulated *PHYA* genes under light conditions (Nieto-Sotelo et al., 1994). Our study found that the Pgβ-glucosidase gene showed an upward trend after 6:00 PM and peaked at 12:00 PM Combined with promoter sequence analysis, the finding revealed certain light-responsive elements such as AE-box, Box 4, GT1-motif, MRE, TCCC-motif, and TCT-motif (**Supplementary Figure 6**). Our results suggest that the gene might be regulated by light and its expression enhanced in dark environments.

In conclusion, we have successfully cloned and characterized the β-glucosidase from *P. grandiflorus*. We verified its function and provided a new perspective for analyzing the biosynthetic pathway of oleanane-type triterpenoid saponins in *P. grandiflorus*. Moreover, we found that the expression of this gene might be regulated by light, which contributes to the further study of its molecular biology. Our findings provide direction for the molecular breeding of *P. grandiflorus* and the improving of the quality of medicinal materials. Future work should investigate the function of β-glucosidase in the plant body and attempt to regulate the conversion of PE to PD in *P. grandiflorus*.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

SX, XS, ZW, LW, and WJ designed the project and/or conducted aspects of the experimental work. XS, ZW, FM, XY, YL, JW, and XG conducted the experiments and the collection of electronic resources. SX, SG, and DP supported this work financially and participated in its planning. XS, SX, and LW wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of China (Grant No. U21A20406), the NSF of Anhui Province (Grant No. 1908085MH268), the Anhui University Collaborative Innovation Project (Grant Nos. GXXT-2019-043 and GXXT-2019-049), the Foundation of Hunan Key Laboratory for Conservation and Utilization of Biological Resources in the Nanyue Mountainous Region (Grant No. NY20K04), and Key Natural Science Research Projects in Anhui Universities (Grant No. KJ2021A0676).

ACKNOWLEDGMENTS

We would like to thank the group of Xiaoya Chen from the CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, for donating the pET-32α(+) vector and *E. coli* BL21(DE3). We also thank Qian Shen from Shanghai Jiaotong University for donating seeds of *N. benthamiana* and the pBI121 vector, and Zhenbang Liu from the Division of Life Sciences and Medicine, University of Science and Technology of China for helping in confocal laser scanning.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.955628/full#supplementary-material

REFERENCES

Ahn, H. J., You, H. J., Park, M. S., Johnston, T. V., Ku, S., and Ji, G. E. (2018). Biocatalysis of platycoside E and platycodin D3 using fungal extracellular β-glucosidase responsible for rapid platycodin D production. *Int. J. Mol. Sci.* 19:2671. doi: 10.3390/ijms19092671

Baglioni, M., Breccia, J. D., and Mazzaferro, L. S. (2021). Peculiarities and systematics of microbial diglycosidases, and their applications in food technology. *Appl. Microbiol. Biotechnol.* 105, 2693–2700. doi: 10.1007/s00253-021-11219-9

Bhatia, Y., Mishra, S., and Bisaria, V. S. (2002). Microbial beta-glucosidases: cloning, properties, and applications. *Crit. Rev. Biotechnol.* 22, 375–407. doi: 10.1080/07388550290789568

Cho, S. Y., Song, C. H., Lee, J. E., Choi, S. H., Ku, S. K., and Park, S. J. (2018). Effects of platycodin D on reflux esophagitis due to modulation of antioxidant defense systems. *Evid. Based Complement. Alternat. Med.* 2018:7918034. doi: 10.1155/2018/7918034

Chuenchor, W., Pengthaisong, S., Robinson, R. C., Yuvaniyama, J., Oonanant, W., Bevan, D. R., et al. (2008). Structural insights into rice BGlu1 beta-glucosidase oligosaccharide hydrolysis and transglycosylation. *J. Mol. Biol.* 377, 1200–1215. doi: 10.1016/j.jmb.2008.01.076

Chuenchor, W., Pengthaisong, S., Robinson, R. C., Yuvaniyama, J., Svasti, J., and Cairns, J. R. (2011). The structural basis of oligosaccharide binding by rice BGlu1 beta-glucosidase. *J. Struct. Biol.* 173, 169–179. doi: 10.1016/j.jsb.2010.09.021

Dehesh, K., Bruce, W. B., and Quail, P. H. (1990). A trans-acting factor that binds to a GT-motif in a phytochrome gene promoter. *Science* 250, 1397–1399. doi: 10.1126/science.2255908

Dziadas, M., Junka, A., and Jeleń, H. (2021). Human saliva-mediated hydrolysis of eugenyl-β-D-glucoside and fluorescein-di-β-d-glucoside in in vivo and in vitro models. *Biomol. Ther.* 11:172. doi: 10.3390/biom11020172

Godse, R., Bawane, H., Tripathi, J., and Kulkarni, R. (2021). Unconventional β-glucosidases: a promising biocatalyst for industrial biotechnology. *Appl. Biochem. Biotechnol.* 193, 2993–3016. doi: 10.1007/s12010-021-03568-y

Guo, R., Meng, Q., Wang, B., and Li, F. (2021). Anti-inflammatory effects of platycodin D on dextran sulfate sodium (DSS) induced colitis and *E. coli* lipopolysaccharide (LPS) induced inflammation. *Int. Immunopharmacol.* 94:107474. doi: 10.1016/j.intimp.2021.107474

Ha, I. J., Ha, Y. W., Kang, M., Lee, J., Park, D., and Kim, Y. S. (2010). Enzymatic transformation of platycosides and one-step separation of platycodin D by high-speed countercurrent chromatography. *J. Sep. Sci.* 33, 1916–1922. doi: 10.1002/jssc.200900842

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* 92, 7090–7094. doi: 10.1073/pnas.92.15.7090

Hiratsuka, K., and Chua, N. H. (1997). Light regulated transcription in higher plants. *J. Plant Res.* 110, 131–139. doi: 10.1007/bf02506852

Horikoshi, S., Saburi, W., Yu, J., Matsuura, H., Cairns, J. R. K., Yao, M., et al. (2022). Substrate specificity of glycoside hydrolase family 1 β-glucosidase AtBGlu42 from Arabidopsis thaliana and its molecular mechanism. *Biosci. Biotechnol. Biochem.* 86, 231–245. doi: 10.1093/bbb/zbab200

Huang, M. Y., Jiang, X. M., Xu, Y. L., Yuan, L. W., Chen, Y. C., Cui, G., et al. (2019). Platycodin D triggers the extracellular release of programed death ligand-1 in lung cancer cells. *Food Chem. Toxicol.* 131:110537. doi: 10.1016/j.fct.2019.05.045

Huang, W., Zhou, H., Yuan, M., Lan, L., Hou, A., and Ji, S. (2021). Comprehensive characterization of the chemical constituents in *Platycodon grandiflorum* by an integrated liquid chromatography-mass spectrometry strategy. *J. Chromatogr. A* 1654:462477. doi: 10.1016/j.chroma.2021.462477

Hudson, M. E., and Quail, P. H. (2003). Identification of promoter motifs involved in the network of phytochrome A-regulated gene expression by combined analysis of genomic sequence and microarray data. *Plant Physiol.* 133, 1605–1616. doi: 10.1104/pp.103.030437

Jenkins, J., Lo Leggio, L., Harris, G., and Pickersgill, R. (1995). Beta-glucosidase, beta-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold beta/alpha architecture and with two conserved glutamates near the carboxy-terminal ends of beta-strands four and seven. *FEBS Lett.* 362, 281–285. doi: 10.1016/0014-5793(95)00252-5

Jeong, E. K., Ha, I. J., Kim, Y. S., and Na, Y. C. (2014). Glycosylated platycosides: identification by enzymatic hydrolysis and structural determination by LC-MS/MS. *J. Sep. Sci.* 37, 61–68. doi: 10.1002/jssc.201300918

Ketudat Cairns, J. R., and Esen, A. (2010). β-glucosidases. *Cell. Mol. Life Sci.* 67, 3389–3405. doi: 10.1007/s00018-010-0399-2

Kil, T. G., Kang, S. H., Kim, T. H., Shin, K. C., and Oh, D. K. (2019). Enzymatic biotransformation of balloon flower root saponins into bioactive platycodin D by deglucosylation with *Caldicellulosiruptor bescii* β-glucosidase. *Int. J. Mol. Sci.* 20:3854. doi: 10.3390/ijms20163854

Kim, J., Kang, S. H., Park, S. G., Yang, T. J., Lee, Y., Kim, O. T., et al. (2020). Whole-genome, transcriptome, and methylome analyses provide insights into the evolution of platycoside biosynthesis in *Platycodon grandiflorus*, a medicinal plant. *Hortic. Res.* 7:112. doi: 10.1038/s41438-020-0329-x

Kwon, J., Lee, H., Kim, N., Lee, J. H., Woo, M. H., Kim, J., et al. (2017). Effect of processing method on platycodin D content in *Platycodon grandiflorum* roots. *Arch. Pharm. Res.* 40, 1087–1093. doi: 10.1007/s12272-017-0946-6

Lee, E. J., Kang, M., and Kim, Y. S. (2012). Platycodin D inhibits lipogenesis through AMPKα-PPARγ2 in 3T3-L1 cells and modulates fat accumulation in obese mice. *Planta Med.* 78, 1536–1542. doi: 10.1055/s-0032-1315147

Li, G., Jiang, Y., Fan, X. J., and Liu, Y. H. (2012a). Molecular cloning and characterization of a novel β-glucosidase with high hydrolyzing ability for soybean isoflavone glycosides and glucose-tolerance from soil metagenomic library. *Bioresour. Technol.* 123, 15–22. doi: 10.1016/j.biortech.2012.07.083

Li, X., Xia, W., Bai, Y., Ma, R., Yang, H., Luo, H., et al. (2018). A novel thermostable GH3 β-glucosidase from talaromyce leycettanus with broad substrate specificity and significant soybean isoflavone glycosides-hydrolyzing capability. *Biomed. Res. Int.* 2018:4794690. doi: 10.1155/2018/4794690

Li, W., Zhao, L. C., Wang, Z., Zheng, Y. N., Liang, J., and Wang, H. (2012b). Response surface methodology to optimize enzymatic preparation of deapio-platycodin D and platycodin D from *Radix Platycodi*. *Int. J. Mol. Sci.* 13, 4089–4100. doi: 10.3390/ijms13044089

Liu, D., Legras, J. L., Zhang, P., Chen, D., and Howell, K. (2021). Diversity and dynamics of fungi during spontaneous fermentations and association with unique aroma profiles in wine. *Int. J. Food Microbiol.* 338:108983. doi: 10.1016/j.ijfoodmicro.2020.108983

Livak, K. J., and Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25, 402–408. doi: 10.1006/meth.2001.1262

Marana, S. R. (2006). Molecular basis of substrate specificity in family 1 glycoside hydrolases. *IUBMB Life* 58, 63–73. doi: 10.1080/15216540600617156

Mehrotra, R., and Panwar, J. (2009). Dimerization of GT element interferes negatively with gene activation. *J. Genet.* 88, 257–260. doi: 10.1007/s12041-009-0037-7

Nieto-Sotelo, J., Ichida, A., and Quail, P. H. (1994). PF1: an A-T hook-containing DNA binding protein from rice that interacts with a functionally defined d(AT)-rich element in the oat phytochrome A3 gene promoter. *Plant Cell* 6, 287–301. doi: 10.1105/tpc.6.2.287

Noh, K. H., Son, J. W., Kim, H. J., and Oh, D. K. (2009). Ginsenoside compound K production from ginseng root extract by a thermostable beta-glycosidase from *Sulfolobus solfataricus*. *Biosci. Biotechnol. Biochem.* 73, 316–321. doi: 10.1271/bbb.80525

Prieto, A., De Eugenio, L., Méndez-Líter, J. A., Nieto-Domínguez, M., Murgiondo, C., Barriuso, J., et al. (2021). Fungal glycosyl hydrolases for sustainable plant biomass valorization: *Talaromyces amestolkiae* as a model fungus. *Int. Microbiol.* 24, 545–558. doi: 10.1007/s10123-021-00202-z

Shi, C., Li, Q., and Zhang, X. (2020). Platycodin D protects human fibroblast cells from premature senescence induced by H_2O_2 through improving mitochondrial biogenesis. *Pharmacology* 105, 598–608. doi: 10.1159/000505593

Shin, K. C., Kim, D. W., Woo, H. S., Oh, D. K., and Kim, Y. S. (2020). Conversion of glycosylated platycoside E to deapiose-xylosylated platycodin D by cytolase PCL5. *Int. J. Mol. Sci.* 21:1207. doi: 10.3390/ijms21041207

Shin, K. C., and Oh, D. K. (2016). Classification of glycosidases that hydrolyze the specific positions and types of sugar moieties in ginsenosides. *Crit. Rev. Biotechnol.* 36, 1036–1049. doi: 10.3109/07388551.2015.1083942

Shin, K.-C., Seo, M.-J., Kim, D.-W., Yeom, S.-J., and Kim, Y.-S. (2019). Characterization of β-glycosidase from *Caldicellulosiruptor owensensis* and its application in the production of platycodin D from balloon flower leaf. *Catalysts* 9:1025. doi: 10.3390/catal9121025

Singh, N., and Kumaria, S. (2020). Molecular cloning and characterization of chalcone synthase gene from Coelogyne ovalis Lindl. and its stress-dependent expression. *Gene* 762:145104. doi: 10.1016/j.gene.2020.145104

Stormo, G. D. (2006). An introduction to recognizing functional domains. *Curr. Protoc. Bioinform.* Chapter 2:Unit 2.1. doi: 10.1002/0471250953.bi0201s15

Su, X., Liu, Y., Han, L., Wang, Z., Cao, M., Wu, L., et al. (2021). A candidate gene identified in converting platycoside E to platycodin D from *Platycodon grandiflorus* by transcriptome and main metabolites analysis. *Sci. Rep.* 11:9810. doi: 10.1038/s41598-021-89294-1

Wang, Y., Che, J., Zhao, H., Tang, J., and Shi, G. (2018). Platycodin D inhibits oxidative stress and apoptosis in H9c2 cardiomyocytes following hypoxia/reoxygenation injury. *Biochem. Biophys. Res. Commun.* 503, 3219–3224. doi: 10.1016/j.bbrc.2018.08.129

Wang, C., Zhang, N., Wang, Z., Qi, Z., Zheng, B., Li, P., et al. (2017). Rapid characterization of chemical constituents of *Platycodon grandiflorum* and its adulterant *Adenophora stricta* by UPLC-QTOF-MS/MS. *J. Mass Spectrom.* 52, 643–656. doi: 10.1002/jms.3967

Wu, J., Yang, G., Zhu, W., Wen, W., Zhang, F., Yuan, J., et al. (2012). Anti-atherosclerotic activity of platycodin D derived from roots of *Platycodon grandiflorum* in human endothelial cells. *Biol. Pharm. Bull.* 35, 1216–1221. doi: 10.1248/bpb.b-y110129

Xu, Z. F., Chye, M. L., Li, H. Y., Xu, F. X., and Yao, K. M. (2003). G-box binding coincides with increased *Solanum melongena* cysteine proteinase expression in senescent fruits and circadian-regulated leaves. *Plant Mol. Biol.* 51, 9–19. doi: 10.1023/a:1020859518877

Yoo, D. S., Choi, Y. H., Cha, M. R., Lee, B. H., Kim, S. J., Yon, G. H., et al. (2011). HPLC-ELSD analysis of 18 platycosides from balloon flower roots (*Platycodi Radix*) sourced from various regions in Korea and geographical clustering of the cultivation areas. *Food Chem.* 129, 645–651. doi: 10.1016/j.foodchem.2011.04.106

Yu, H., Liu, M., Yin, M., Shan, T., Peng, H., Wang, J., et al. (2021). Transcriptome analysis identifies putative genes involved in triterpenoid biosynthesis in *Platycodon grandiflorus*. *Planta* 254:34. doi: 10.1007/s00425-021-03677-2

Zada, N. S., Belduz, A. O., Güler, H. I., Khan, A., Sahinkaya, M., Kaçıran, A., et al. (2021a). Cloning, expression, biochemical characterization, and molecular docking studies of a novel glucose tolerant β-glucosidase from Saccharomonospora sp. NB11. *Enzym. Microb. Technol.* 148:109799. doi: 10.1016/j.enzmictec.2021.109799

Zada, N. S., Belduz, A. O., Güler, H. I., Sahinkaya, M., Khan, S. I., Saba, M., et al. (2021b). Cloning, biochemical characterization and molecular docking of novel thermostable β-glucosidase BglA9 from *Anoxybacillus ayderensis* A9 and its application in de-glycosylation of Polydatin. *Int. J. Biol. Macromol.* 193, 1898–1909. doi: 10.1016/j.ijbiomac.2021.11.021

Zechel, D. L., and Withers, S. G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* 33, 11–18. doi: 10.1021/ar970172+

Zhang, L., Wang, Y., Yang, D., Zhang, C., Zhang, N., Li, M., et al. (2015). *Platycodon grandiflorus*: an ethnopharmacological, phytochemical and pharmacological review. *J. Ethnopharmacol.* 164, 147–161. doi: 10.1016/j.jep.2015.01.052

Zhang, Y., Zhang, Y., Liang, J., Kuang, H. X., and Xia, Y. G. (2022). Exploring the effects of different processing techniques on the composition and biological activity of *Platycodon grandiflorus* (Jacq.) A.DC. by metabonomics and pharmacologic design. *J. Ethnopharmacol.* 289:114991. doi: 10.1016/j.jep.2022.114991

Zhang, Z., Zhao, M., Zheng, W., and Liu, Y. (2017). Platycodin D, a triterpenoid saponin from *Platycodon grandiflorum*, suppresses the growth and invasion of human oral squamous cell carcinoma cells via the NF-κB pathway. *J. Biochem. Mol. Toxicol.* 31:e21934. doi: 10.1002/jbt.21934

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 15 July 2022
DOI 10.3389/fpls.2022.951605

OPEN ACCESS

EDITED BY
Inna Khozin-Goldberg,
Ben-Gurion University of the Negev,
Israel

REVIEWED BY
Ralf Welsch,
University of Freiburg,
Germany
Shan Lu,
Nanjing University,
China

*CORRESPONDENCE
Yongsheng Tian
ytian_saas@163.com
Rihe Peng
rpeng_saas@163.com
Quanhong Yao
qhyao_saas@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Plant Systems and Synthetic Biology, a section of the journal Frontiers in Plant Science

RECEIVED 24 May 2022
ACCEPTED 28 June 2022
PUBLISHED 15 July 2022

CITATION
Li Z, Gao J, Wang B, Xu J, Fu X, Han H, Wang L, Zhang W, Deng Y, Wang Y, Gong Z, Tian Y, Peng R and Yao Q (2022) Rice carotenoid biofortification and yield improvement conferred by endosperm-specific overexpression of *OsGLK1*. *Front. Plant Sci.* 13:951605. doi: 10.3389/fpls.2022.951605

Rice carotenoid biofortification and yield improvement conferred by endosperm-specific overexpression of *OsGLK1*

Zhenjun Li†, Jianjie Gao†, Bo Wang†, Jing Xu, Xiaoyan Fu, Hongjuan Han, Lijuan Wang, Wenhui Zhang, Yongdong Deng, Yu Wang, Zehao Gong, Yongsheng Tian*, Rihe Peng* and Quanhong Yao*

Shanghai Key Laboratory of Agricultural Genetics and Breeding, Agro-Biotechnology Research Institute, Shanghai Academy of Agricultural Sciences, Shanghai, China

Carotenoids, indispensable isoprenoid phytonutrients, are synthesized in plastids and are known to be deficient in rice endosperm. Many studies, involving transgenic manipulations of carotenoid biosynthetic genes, have been performed to obtain carotenoid-enriched rice grains. Nuclear-encoded GOLDEN2-LIKE (GLK) transcription factors play important roles in the regulation of plastid and thylakoid grana development. Here, we show that endosperm-specific overexpression of rice *GLK1* gene (*OsGLK1*) leads to enhanced carotenoid production, increased grain yield, but deteriorated grain quality in rice. Subsequently, we performed the bioengineering of carotenoids biosynthesis in rice endosperm by introducing other three carotenogenic genes, *tHMG1*, *ZmPSY1*, and *PaCrtI*, which encode the enzymes truncated 3-hydroxy-3-methylglutaryl-CoA reductase, phytoene synthase, and phytoene desaturase, respectively. Transgenic overexpression of all four genes (*OsGLK1*, *tHMG1*, *ZmPSY1*, and *PaCrtI*) driven by rice endosperm-specific promoter *GluB-1* established a mini carotenoid biosynthetic pathway in the endosperm and exerted a roughly multiplicative effect on the carotenoid accumulation as compared with the overexpression of only three genes (*tHMG1*, *ZmPSY1*, and *PaCrtI*). In addition, the yield enhancement and quality reduction traits were also present in the transgenic rice overexpressing the selected four genes. Our results revealed that *OsGLK1* confers favorable characters in rice endosperm and could help to refine strategies for the carotenoid and other plastid-synthesized micronutrient fortification in bioengineered plants.

KEYWORDS

grain yield, *OsGLK1*, plastids, rice endosperm, carotenoids

Introduction

Carotenoids are a diverse group of hydrophobic isoprenoid pigments ubiquitously synthesized by plants, algae, fungi, and bacteria (Sun et al., 2018). Carotenoids play critical roles in light-harvesting and protection of the photosynthetic apparatus against damage by high light and/or temperature stress (Nisar et al., 2015). Carotenoids provide precursors

for the biosynthesis of plant hormones and serve as important contributors to visual appeal of ornamental plants and nutritional qualities of fruits and vegetables (Cazzonelli and Pogson, 2010; Ruiz-Sola and Rodríguez-Concepción, 2012).

Most cereal crops contain only trace levels of dietary carotenoids. Carotenoids are currently under intense scrutiny regarding their potential to reduce the risks of some chronic diseases and prevent retinol (vitamin A) deficiency. Thus, a great deal of research effort has been made to modify the types and levels of carotenoids in economically important crops (Paine et al., 2005). The carotenoid biosynthesis at the molecular level in higher plants has been well elucidated. Plant carotenoids are isopentenyl diphosphate (IPP)-derived molecules, and two pathways independently contribute to the production of IPP in plants: the cytosolic mevalonic acid (MVA) pathway and the plastidic methylerythritol 4-phosphate (MEP) pathway (Hemmerlin et al., 2012). In plant cells, plastidic MEP pathway is the major pathway for carotenoid biosynthesis and storage (Lopez-Juez and Pyke, 2005). Many studies have been focused on increasing metabolic flux in plastids by overexpression of one or more crucial enzymes within MEP pathway for carotenoid accumulation (Shewmaker et al., 1999; Zhai et al., 2016; Paul et al., 2017). However, very few genes that are not from the two pathways have been well characterized for their influence on carotenoid production in plants.

The endosperm is the largest component of the whole cereal grains, and serves as major food staple, but it is deficient in many nutritionally valuable biochemical compounds such as vitamins, minerals, and carotenoids (Zhu et al., 2007, 2008). Previous studies have shown that some carotenoids are not essential but beneficial to health, whereas some, such as pro-vitamin A carotenoids, have been considered as essential nutrients because they are unable to be synthesized *de novo* by humans and must be acquired through the diet (Fraser and Bramley, 2004; Bai et al., 2011). Rice (*Oryza sativa L.*) is a major cereal food crop in most of the developing countries. Due to the absence of carotenoids in rice endosperm, the consumption of rice as an integral part of diet is often accompanied by vitamin A deficiency (Underwood and Arthur, 1996; Farré et al., 2010).

Biofortification is an economically effective process by which crops are bred in a way that increases the production of micronutrients or phytonutrients *via* genetic engineering (Bouis and Saltzman, 2017). Biofortified crops enriched with phytonutrients, such as flavonoid-enriched tomato (Le Gall et al., 2003), astaxanthin-enriched rice (Zhu et al., 2018), vitamin B_6-enriched rice (Mangel et al., 2019), and anthocyanin-enriched wheat (Sharma et al., 2020), may have the potential to bring great benefits to human health (Hirschi, 2009; De Steur et al., 2015; Fang et al., 2018).

The availability of a large number of carotenogenic genes makes it possible to modify and engineer the carotenoid biosynthetic pathways in host plants. Combinatorial coexpression by a single expression vector harboring all the target genes is commonly regarded as a time-saving and highly efficient plant transformation strategy in comparison with cotransformation of multiple constructs and transgene stacking by crossing (Fang et al., 2018).

The overexpression of carotenoid biosynthetic genes can boost carotenoid accumulation in staple crops if upstream pathways could supply sufficient isoprenoid precursors (Rodríguez-Concepción, 2010). Downstream contributors of carotenoid metabolism and sequestration serve to deplete the carotenoid pool and drive the reaction forward (Auldridge et al., 2006; Campbell et al., 2010). The first committed step of carotenoid biosynthesis is initiated by the condensation of two geranylgeranyl diphosphate (GGPP) molecules. Phytoene synthase (PSY) is the rate-limiting enzyme that catalyzes this reaction in the carotenoid pathway (Cazzonelli and Pogson, 2010). There is also limited flux in the subsequent desaturation reaction, which eventually results in the production of lycopene (Nisar et al., 2015). In order to promote phytoene-to-lycopene conversion, the bacterial phytoene desaturase enzyme *CrtI* was coexpressed with *PSY* in rice endosperm and finally contributed to the accumulation of β-carotene (Paine et al., 2005). Although the expression of these two enzymes increases metabolic flux in the early part of the carotenoid biosynthesis pathway, the exhaustion of the precursor pool remains an unresolved problem (Ye et al., 2000; Paine et al., 2005; Schaub et al., 2005).

3-Hydroxy-3-methylglutaryl coenzyme A reductase (HMGR) governs the MVA pathway-derived isoprenoids and plays a crucial role in isoprenoid biosynthesis. Transgenic tomatoes overexpressing plant *HMGR* without also overexpressing *PSY* and *CrtI* genes show IPP accumulation and no carotenoid accumulation (Enfissi et al., 2005). Although *HMGR* localizes in the cytosol and belongs to the MVA pathway, the overexpression of *HMGR* could significantly boost the production of IPP which may ultimately contribute to the downstream carotenoid biosynthesis. To address this challenge, we heterologously expressed the maize (*Zea mays*) *PSY* (*ZmPSY1*) and *Pantoea ananatis CrtI* (*PaCrtI*) genes along with *tHMG1*, which encodes truncated HMGR (N-terminal 554-amino acid-deletion) from *Saccharomyces cerevisiae*, specifically in the endosperm to boost flux through the MVA pathway, which generates carotenoid precursors. It was found that the combined expression of *tHMG1*, *ZmPSY1*, and *PaCrtI* under the control of endosperm-specific promoter *GluB-1* significantly promote carotenoid accumulation in rice endosperm, confirming that the participation of *tHMG1* in supply of isoprenoid precursors created a metabolic sink for the sustainable production of carotenoids (Tian et al., 2019).

Abbreviations: DW, Dry weight; HPLC, High-Performance Liquid Chromatography; LC/MS, Liquid chromatography-mass spectrometry; MEP, Methylerythritol 4-phosphate pathway; MVA, Mevalonic acid pathway; PTDS, PCR-based two-step DNA synthesis; SD, Standard deviation; SEM, Scanning electron microscopy; TEM, Transmission electron microscopy; UHPLC-QTOFMS, Ultra-High-Performance Liquid Tandem Chromatography Quadrupole Time of Flight Mass Spectrometry.

De novo carotenoids are synthesized within almost all types of plastids, most importantly, chloroplasts and chromoplasts, in leaves, fruits, flowers, roots, and seeds (Yuan et al., 2015; Sun et al., 2018). Plastid identity is a key carotenoid-defining parameter which plays essential roles in governing carotenoid stability and composition. In chloroplasts, the majority of carotenoids are localized in the thylakoid membranes and plastoglobules, which are identified as the sites of carotenoid storage (Chaudhary et al., 2009). The *GOLDEN2-LIKE* (*GLK*) genes encode transcription factors that regulate plastid development and chlorophyll levels in diverse species, namely in the monocot maize and in the eudicot *Arabidopsis*. Defects in plastid development and chlorophyll biosynthesis usually result in compromised carotenoid biosynthesis (Langdale and Kidner, 1994; Hall et al., 1998; Rossini et al., 2001; Fitter et al., 2002). Previous studies showed that overexpression of *GLK*s could significantly promote chlorophyll biosynthesis and increase chloroplast number and thylakoid grana stacks in plant (Fitter et al., 2002; Yasumura et al., 2005; Nguyen et al., 2014; Chen et al., 2016).

Therefore, we expressed the *OsGLK1* gene (genotype G) specifically in the rice endosperm to investigate its roles in carotenoid production. As a comparison, we also create another construct containing four genes encoding *OsGLK1*, *tHMG1*, *ZmPSY1*, and *PaCrtI* (genotype GHPC) to investigate their additive effects on the biofortification of carotenoids in rice endosperm (Figure 1A). It was found that *OsGLK1* significantly enhanced carotenoid production in rice endosperm and increased grain yield. However, the grain quality was adversely affected. Meanwhile, the combined overexpression of *OsGLK1*, *tHMG1*, *ZmPSY1*, and *PaCrtI* led to multiplicative effects on the carotenoid accumulation in rice endosperm compared with previous three-gene combination (*tHMG1*, *ZmPSY1*, and *PaCrtI*, genotype HPC) from our lab (Tian et al., 2019). Our data presented a conceptual and mechanistic basis to generate "New Golden Rice" and to update knowledge on the characterization of the carotenoid biosynthesis pathway.

Materials and methods

Vector construction

The transgenes, based on rice *GLK1* (GenBank: AAK50393), *S. cerevisiae* truncated *HMGR* (GenBank: AJS96703), maize *PSY1* (GenBank: AAB60314), and *P. ananatis CrtI* (GenBank: AHG94990) were chemically synthesized through PCR-based two-step DNA synthesis method (PTDS; Xiong et al., 2006). The codons of these genes were optimized in accordance with the codon usage frequencies of rice. Primers used for chemical synthesis of the four genes are listed in Supplementary Table S1. Then, the two gene expression cassettes (*G:OsGLK1* and *G:OsGLK1-G:tHMG1-G:ZmPSY1-G:PaCrtI*) were constructed by inserting the ORFs between an endosperm-specific promoter *GluB-1* and terminator using the PAGE-mediated overlap extension PCR method (Peng et al., 2014). The original expression vector for rice transformation was pCAMBIA1301 which was modified through the introduction of new multiple cloning sites by our lab to yield the binary vector pYP694 (Peng et al., 2014). Then, *pYP69-G:OsGLK1* (genotype G) was generated by inserting the *G:OsGLK1* expression cassette into pYP694 at the *Eco* RI and *Hin*d III sites. Then the *pYP69-G:tHMG1-G:ZmPSY1-G:PaCrtI* vector (genotype HPC) provided by Tian et al. (2019) was used as the donor vector for the introduction of *G:OsGLK1* to produce the four-gene construct *pYP69-G:OsGLK1-G:tHMG1-G:ZmPSY1-G:PaCrtI* (genotype GHPC). The detailed method was mentioned as previously described (Tian et al., 2019).

Plant materials, strains, and chemicals

The rice conventional variety ZH11 was used for transformation of the two constructs described above. The *Agrobacterium tumefaciens* strain GV3101 was used for rice transformation. The field experiments were performed during the growing season in Shanghai and Hainan, China. Commercially available carotenoid standards, including α-carotene, β-carotene, lycopene, zeaxanthin, lutein, neoxanthin, and astaxanthin were purchased from Sigma Chemical Company (St. Louis, MO, United States).

Carotenoid extraction and quantitative analysis by HPLC

Carotenoids were extracted from endosperms of different genotypes. The extraction method was described previously (Zhu et al., 2018). The HPLC analysis was performed using an Agilent 1,100 HPLC system (Agilent Technologies, CA, United States). The chromatographic separation was achieved with Agilent LC-C_{18} (4.6 mm × 250 mm, 5 μm) column and a gradient system with the mobile phase consisting of solvent A (acetonitrile) and solvent B (isopropanol) at a flow rate of 1 ml/min. The linear gradient was as follows: 0–10.0 min, 100% A; 10.0–13.0 min, 60% A; 13.0–20.0 min, 50% A; 20.0–30.0 min, 100% A. The UV spectra were recorded with a diode-array detector at 450 nm. Identification and peak assignment of each carotenoid component were primarily based on the comparison of their retention time and UV–visible spectrum data with that of standards.

Metabolite profiling

Endosperms of mature ZH11 and G1 seeds were subjected to metabolomic analysis in our study. Metabolomic analysis of rice endosperms was performed using UHPLC-QTOFMS technology (Ultra-High-Performance Liquid Tandem Chromatography Quadrupole Time of Flight Mass Spectrometry). Three biological replicates were performed for each genotype, with each replicate being a pool of 5 g endosperms. The protocols for sample preparation have been described in detail (Wu et al., 2020;

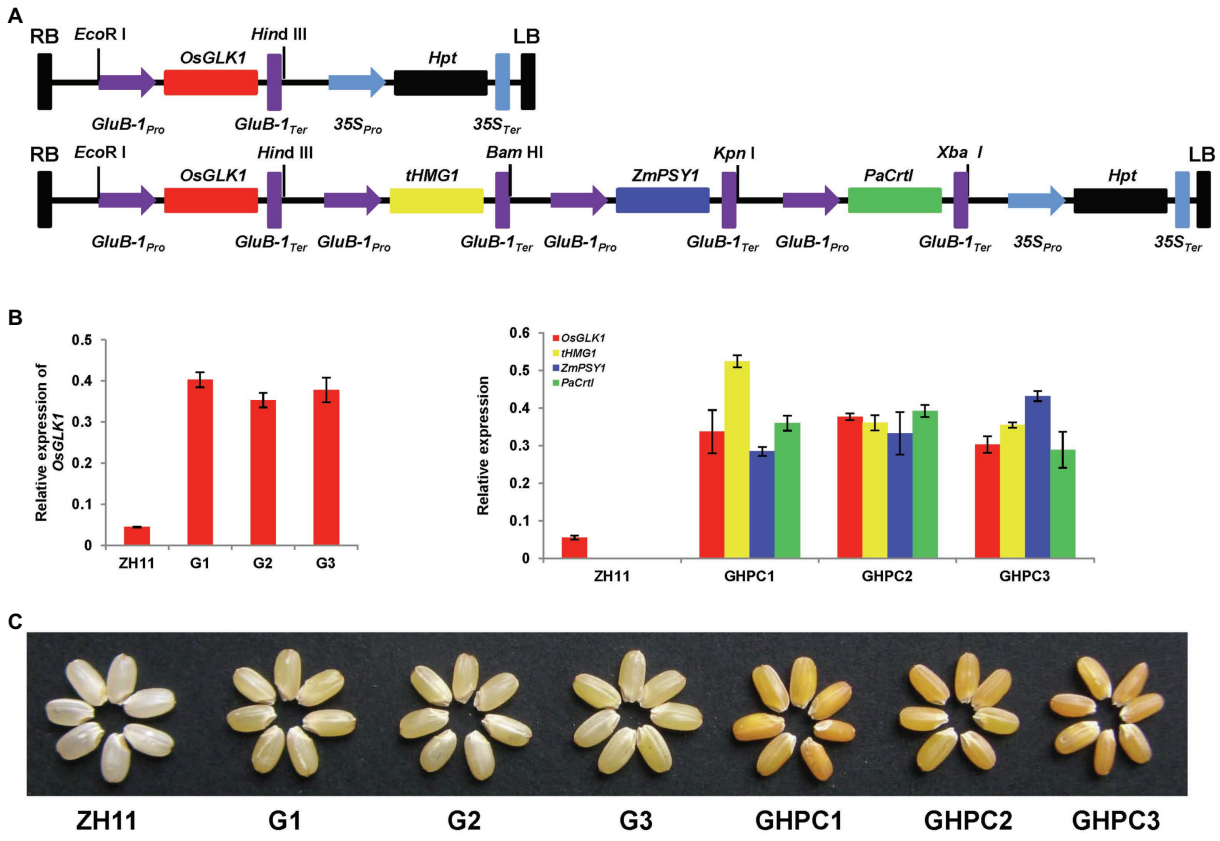


FIGURE 1
Engineering the biosynthesis of carotenoid in rice endosperm. **(A)** Schematic representation of the gene cassettes in the two plasmids used for rice transformation. Upper: vector harboring *OsGLK1*; Lower: vector harboring four genes (*OsGLK1*, *tHMG1*, *ZmPSY1*, and *PaCrtI*). Target genes were all driven by endosperm-specific promoter *GluB-1*. **(B)** Expression levels of the transgenes in homologous T3 lines of G and GHPC rice endosperms. RNA samples were prepared from endosperms of developing seeds (9 days after pollination) of the homologous T3 lines. The expression levels of the transgenes were normalized to the expression of reference gene *OsActin*. Values represent means of three replicates±SD. **(C)** Representative phenotype of transgenic rice seeds showing altered color due to carotenoid accumulation. Two different endosperm phenotypes were observed. The light-yellow endosperm expressed *OsGLK1* (genotype G), and the orange-yellow endosperm coexpressed *OsGLK1*, *tHMG1*, *ZmPSY1*, and *PaCrtI* (genotype GHPC).

Liu et al., 2021). The UHPLC separation was carried out using a Waters ACQUITY UPLC HSS T3 column (100 × 2.1 mm, 1.8 μm). The AB Sciex QTOF mass spectrometer was used to acquire MS/MS spectra. Heatmaps and hierarchical cluster analysis were generated by Multi-Experiment Viewer (version 4.8.1).

Transmission electron microscopy

The transmission electron microscopy observation was performed according to a protocol previously published (Luo et al., 2013) with a slight modification. Rice samples were first fixed overnight (4°C) in 2.5% (v/v) glutaraldehyde (pH 7.2), postfixed in 1% osmium tetroxide at 4°C for 1 h, followed by dehydration through an ethyl alcohol series. After dehydration, samples were embedded in epon araldite (resin) and polymerization was conducted at 40°C for 24 h. Samples were then transferred to fresh resin and hardened under nitrogen air at 60°C for 2 days, followed by sectioning of samples using Leica EM UC7 ultramicrotome (Wetzlar). Sections were stained with 5% uranyl acetate and imaged with a H7100FA transmission electron microscope (Hitachi).

Scanning electron microscopy

Scanning electron microscopy observation was performed as described in the previous studies (Imai et al., 2006). Grain samples were transversely cut by a knife and coated with gold by E-100 ion sputter. Specimens were observed using a scanning electron microscope (SEM, VEGA3, TESCAN, Czech Republic). All procedures were conducted according to the manufacturer's protocol.

Expression analysis

Total RNA was extracted from 0.1 g of immature milky-stage endosperm (9 days after pollination) using an RNeasy Mini kit

(Qiagen, Germany) according to the manufacturer's instructions. All isolated RNA samples were treated with RNase-free DNase I (Promega, United States). Complementary DNA (cDNA) was synthesized from total RNA (2 μg) using a Quantitect reverse transcription kit (Qiagen) following the supplier's protocol. The expression of the four transgenes (*OsGLK1*, *tHMG1*, *ZmPSY1*, and *PaCrtI*) and 11 endogenous carotenogenic genes (*OsPSY*, *OsPDS*, *OsZISO*, *OsZDS*, *OsCRTISO*, *OsLCYE*, *OsLCYB*, *OsBHY*, *OsEHY*, *OsZEP*, and *OsVED*) were investigated by qRT-PCR using SYBR Premix Ex Taq II (Takara, Japan) according to the supplier's instructions. The expression of target genes were normalized to the reference gene *OsActin* (Li et al., 2019). PCR conditions for amplification of the four transgenes were as follows: 4 min at 95°C, followed by 35 cycles of 10 s at 95°C, 15 s at 59°C (54°C for the endogenous carotenogenic genes), 20 s at 72°C. Data are shown as means ± SD (three biological replicates). The information of primers used is shown in Supplementary Table S2. The loci of the 11 endogenous carotenogenic genes in The Rice Annotation Project Database[1] were as follows: *OsPSY*, Os12g0626400; *OsPDS*, Os03g0184000; *OsZISO*, Os12g0405200; *OsZDS*, Os07g0204900; *OsCRTISO*, Os11g0572700; *OsLCYB*, Os02g0190600; *OsLCYE*, Os01g0581300; *OsBHY*, Os10g0533500; *OsEHY*, Os03g0125100; *OsZEP*, Os04g0448900; and *OsVDE*, Os04g0379700.

Trait measurement

Fully filled grains were used for measuring the 1,000-grain weight. The grains with chalkiness were counted, and the percentage of chalky grains was calculated as the chalkiness percentage. For chalkiness degree, 50 grains with chalkiness were randomly selected, and the ratio of the chalkiness area to the whole kernel square for each grain was evaluated. In order to easily distinguish the chalkiness regions from other regions, the images of rice kernels were captured using a digital camera (Canon) and processed by Image J (version 1.8.0). The chalkiness degree was defined as the ratio of the chalkiness area to the whole kernel square. The graphical plots were prepared using GraphPad Prism software (Version 8.3.0).

Results

Generation of carotenoids in rice

Agrobacterium-mediated transformation of ZH11 was carried out to create transgenic plants. Nine independent G and 6 independent GHPC transgenic lines were obtained. Homozygous transgenic rice lines, harboring a single copy of targeted construct (Figure 1A), were selected for the subsequent studies. Subsequently, qRT-PCR analysis was performed to examine the transcript levels of the four transgenes in the endosperm of developing seeds. Since the transcript levels of all transgenes were greatly enhanced in the endosperm, we designated these lines as G1, G2, G3, GHPC1, GHPC2, and GHPC3 lines (Figure 1B). All transgenic rice produced light-yellow and orange-yellow endosperms for G and GHPC lines, respectively (Figure 1C).

Composition of carotenoids and their contents in transgenic lines

Carotenoid concentrations in rice endosperms were measured by HPLC to ascertain the effect of *OsGLK1* on pigment accumulation. Seven major carotenoids (lycopene, α-carotene, β-carotene, lutein, zeaxanthin, neoxanthin, and astaxanthin) were quantified. Quantitative analysis of the contents of these components showed that overexpression of *OsGLK1* could significantly increase carotenoid production in rice endosperm. The total carotenoid content was 2.13, 2.37, and 2.32 μg/g dry weight (DW) in lines G1, G2, and G3, respectively. It needs to be noticed that the content of β-carotene increased to a greater extent in all G lines compared with the production of other carotenoid components (Table 1; Supplementary Figure S1). Metabolite profiling of rice endosperms of line G1 and ZH11 identified 89 metabolites with altered levels, only a small number of which showed changes greater than 2 folds, suggesting that critical metabolic activities were unaffected in *OsGLK1* overexpression plants (Supplementary Table S3; Supplementary Figure S2).

The GHPC lines produced the highest levels of carotenoids (39.36, 40.06, and 36.81 μg/g DW for lines GHPC1, GHPC2, and GHPC3, respectively; Table 1; Supplementary Figure S1). This revealed an increase in total carotenoids of nearly 3-fold compared to the previously reported HPC rice endosperms (up to 14.45 μg/g DW) from our lab (Tian et al., 2019), supporting the notion that introduction of the *OsGLK1* gene has a big impact on carotenoids production in rice. The total carotenoid contents ranged from 2.13 to 2.37 μg/g DW and 13.71–14.45 μg/g DW in G and HPC lines, respectively. However, the GHPC lines produced more than 40 μg/g DW of total carotenoids. These results indicate that *OsGLK1* might have roughly multiplicative, rather than additive effects on the carotenoids accumulation in our study. In addition, qRT-PCR analysis was performed to investigate the expression levels of *tHMG1*, *ZmPSY1*, and *PaCrtI* in HPC lines. The results showed that the transcript levels of the three genes in GHPC lines are comparable with those in HPC lines (Figure 1B; Supplementary Figure S3), which further justified the extraordinary effect of *OsGLK1* on carotenoid accumulation. Transmission electron microscopy revealed that a few electron-dense plastoglobuli (which may contain carotenoids) were visible inside some plastids of 15 DAP (days after pollination) endosperm in the genotypes G and GHPC. However, we can barely observe plastoglobuli in the ZH11 and HPC endosperms (Figure 2A; Supplementary Figure S4A). The formation of plastoglobuli in the plastids of transgenic rice

1 https://rapdb.dna.affrc.go.jp/

TABLE 1 Carotenoid content and composition of mature rice endosperms from different genotypes.

Composition (μg/g DW)	Genotype						
	ZH11	G1	G2	G3	GHPC1	GHPC2	GHPC3
α-Carotene	0 ± 0	0.247 ± 0.057***	0.237 ± 0.044***	0.243 ± 0.048***	9.472 ± 0.112***	9.547 ± 0.042***	8.839 ± 0.41***
β-Carotene	0.009 ± 0.001	1.474 ± 0.11***	1.48 ± 0.086***	1.572 ± 0.016***	28.158 ± 0.283***	28.774 ± 0.52***	26.227 ± 0.634***
Lycopene	0 ± 0	0.105 ± 0.009***	0.222 ± 0.012***	0.209 ± 0.018***	0.796 ± 0.001***	0.798 ± 0.002***	0.795 ± 0.001***
Zeaxanthin	0.005 ± 0.001	0.072 ± 0.003***	0.093 ± 0.006***	0.072 ± 0.007***	0.295 ± 0.003***	0.299 ± 0.016***	0.276 ± 0.013***
Lutein	0.025 ± 0.002	0.231 ± 0.01***	0.342 ± 0.007***	0.228 ± 0.006***	0.635 ± 0.004***	0.64 ± 0.023***	0.673 ± 0.027***
Total carotenoid	0.039 ± 0.004	2.128 ± 0.089***	2.373 ± 0.086***	2.323 ± 0.022***	39.356 ± 0.693***	40.058 ± 0.628***	36.81 ± 0.538***

Data are presented as means ± SD of three biological replicates. *p*-values produced by two-tailed Student's *t*-test. DW, dry weight. ***$p < 0.001$.

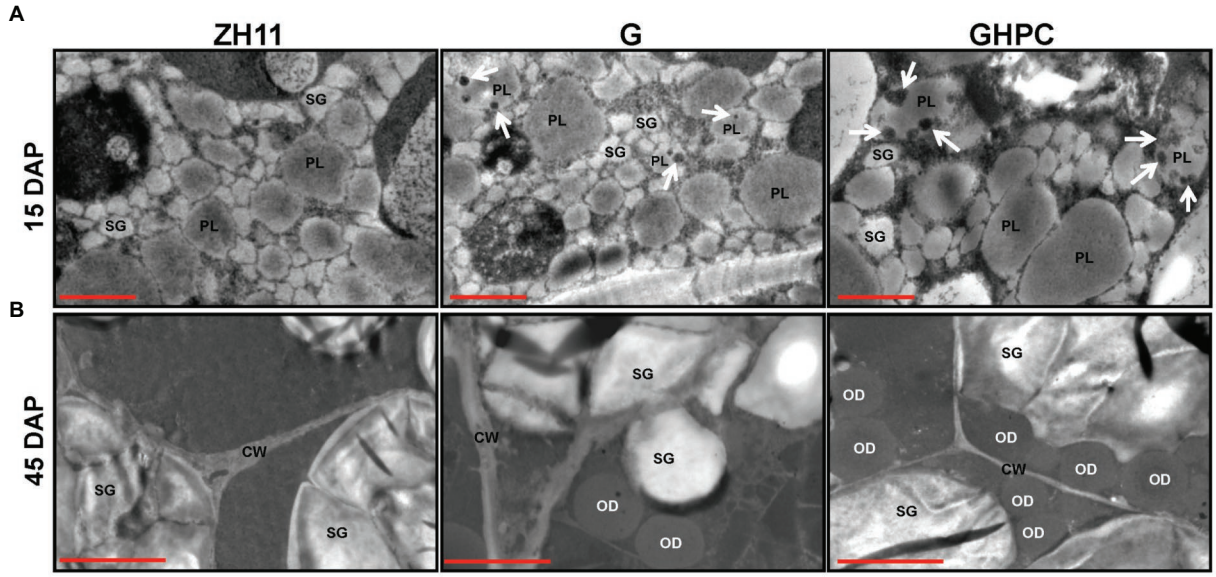


FIGURE 2
Transmission electron micrographs of rice endosperm. **(A)** The endosperm cells (15 DAP) of ZH11, G, and GHPC genotypes are shown. Scale bar=1μm. **(B)** The endosperm cells (45 DAP) of ZH11, G, and GHPC genotypes are shown. Scale bar=2μm. DAP, days after pollination. White arrows indicate the electron-dense plastoglobuli in plastids (PL). SG, starch grain; OD, oil droplet; CW, cell walls.

endosperm might be, at least in part, due to the overexpression of *OsGLK1* in our study. In addition, TEM of G, HPC, and GHPC grains revealed numerous oil droplets (also called spherosomes) in the 45 DAP endosperm, while no oil droplets could be observed in ZH11 endosperm (Figure 2B; Supplementary Figure S4B). It was previously reported that the elevation of the predominant carotenoid is consistent with elevated plastid numbers (Nguyen et al., 2014). The multiplicative effect on the burst production of carotenoids imposed by *OsGLK1* might be due to the improved thylakoid membrane biogenesis and plastid structure and number. *GluB-1* promoter is commonly used for foreign gene expression only in the endosperm area (Washida et al., 1999). TEM was also employed to monitor the chloroplast development in 15 DAP seed coats of each genotype. As expected, no significant differences among ZH11, G and GHPC genotypes regarding the seed coat color were detected (Supplementary Figure S5A). The chloroplasts in seed coats from each genotype developed normally with no obvious phenotypic difference (Supplementary Figures S5B,C).

Expression of endogenous carotenoid biosynthetic genes

In accordance with the previous literatures (Hirschberg, 2001; Vranová et al., 2013; Giuliano, 2014; Nisar et al., 2015), we summarized the expanded carotenoid biosynthetic pathway in plants (Figure 3A). In order to examine how differences in the carotenoid complement may be linked to differential changes in endogenous gene expression, qRT-PCR was used to investigate the accumulation of transcripts corresponding to the 11 genes of the carotenoid biosynthetic pathway in transgenic rice. Two lines of

G and GHPC genotypes, respectively, were selected for the gene expression analysis. All of the analyzed endogenous carotenogenic genes were found to be expressed, some at relatively low levels, in the endosperm of ZH11 seeds. The mRNA levels of most endogenous genes were not greatly interfered by the overexpression of transgenes in both G and GHPC lines (Figure 3B). It is noteworthy that the transcript levels of some endogenous genes (*OsPDS*, *OsZDS*, *OsLCYB*, and *OsVDE*) were slightly up-regulated in the transgenic lines, probably due to positive feedback regulation in response to the overexpression of transgenes, which is common in the regulation of cellular processes (Schaub et al., 2005; Mitrophanov and Groisman, 2008) and also was observed in some carotenogenic transgenic plants (Huang et al., 2013; Bai et al., 2016). The enhanced expression level of endogenous carotenogenic gene *OsLCYB* might play a part in the higher accumulation of β-carotene in G and GHPC endosperms. In another branch, the expression level of *OsLCYE* was relatively low in all genotypes, but obviously sufficient for the enzyme activity to catalyze lycopene into α-carotene (Figure 3B).

Increased grain yield and deteriorated grain quality

Rice is one of the most important food crops in the world. In order to provide a comprehensive assessment for the application of our transgenic plants, important developmental and economic traits, such as grain yield and quality, were investigated in our study. A phenotypic evaluation of G, GHPC and nontransgenic ZH11 plants revealed no major differences at the vegetative growth stage. To evaluate yield potential of G and GHPC rice lines in field conditions, we performed experiments at two different locations (Beijing and Hainan) that have distinctive climates and day lengths. Hainan is warmer than Shanghai throughout most of the season. Table 2 shows that, at both field sites, plant vegetative traits including height, tiller and panicle number, panicle length, and grain number of per panicle were not significantly changed in both G and GHPC lines compared to those in ZH11, whereas the seed setting rates of G and GHPC lines were consistently greater than those of ZH11 with differences being significant at both sites. Meanwhile, the 1,000-grain weight of G and GHPC genotypes was both slightly decreased compared with that of ZH11 (Table 2). Despite the reduced 1,000-grain weight, higher seed-setting percentage, an ~12% increase in seed-setting rate in G and GHPC lines, translated into a 17.6%–23.9% increase in grain yield (Table 2; Figure 4).

Surprisingly, the percentage and degree of chalkiness in G and GHPC were seriously increased compared with those in ZH11 (Figures 5A,B). The above results suggest that *OsGLK1* and/or the carotenoid overproduction might be involved in chalky endosperm formation. In the white light background, we can easily observe that the ZH11 grains displayed lower chalkiness compared with the G and GHPC grains (Figures 6A,B). The results of scanning electron microscopy revealed that the starch granules in the crystal areas of ZH11, G and GHPC grains were regular-shaped and closely arranged (Figure 6C). While, the starch granules in the chalky area of G lines were spherical, loosely packed and irregularly polyhedron-shaped (Figure 6D). The endosperms of GHPC lines consist of relatively smaller spherical starch granules (Figure 6D). In addition, waxy starch granule phenotype was observed in the endosperm of GHPC grains which might be due to the extremely high accumulation of carotenoids (Figure 6C).

Discussion

The generation of Golden Rice 2 variety (Paine et al., 2005) justified the effectiveness of carotenoid biofortification by introducing functional genes involved in carotenoid biosynthetic pathway. For this purpose, *tHMG1*, *ZmPSY1*, and *PaCrtI* were chemically synthesized with codon optimization for rice based on the protein sequences of truncated *S. cerevisiae HMGR*, maize *PSY1*, and *P. ananatis CrtI*, respectively (Tian et al., 2019). The three gene expression cassette controlled by endosperm-specific promoter *GluB-1* from the rice globulin gene was transformed to ZH11 and greatly enhanced carotenoid production (Tian et al., 2019).

Carotenoids are synthesized in chloroplasts, chromoplasts, and other plastids in plants (Cazzonelli and Pogson, 2010; Ruiz-Sola and Rodríguez-Concepción, 2012; Sun et al., 2018). Due to the positive effects of *GLK* genes on plastid and thylakoid grana stacks development (Rossini et al., 2001; Fitter et al., 2002), it is tempting to reason that introducing a rice *GLK* transcription factor gene (*OsGLK1*), along with those for *tHMG1*, *ZmPSY1*, and *PaCrtI* may contribute to effective biofortification of carotenoids in rice endosperm. Enhancement of carotenoid production in G and GHPC lines indicated that *OsGLK1* has the potential to manipulate the plastid sink strength and enhance the availability of storage structures in rice grains. The overexpression of *OsGLK1* could be used as a "push" strategy on increasing metabolic flux in plastids to enlarge the carotenoid pool size.

Both the Golden Rice and Golden Rice 2 lines accumulated predominantly β-carotene, but phytoene, the immediate product of PSY, was completely exhausted (Ye et al., 2000; Paine et al., 2005). The buildup of β-carotene is likely to affect the resource balance in the pathway and block further conversion, especially of downstream metabolites. In other words, biosynthetic and degradative reactions and subcellular environments for deposition and sequestration within and outside of plastids have the potential to affect the final carotenoid composition. Therefore, a further enhancement of carotenoid production in rice may not only require the supply of more isoprenoid precursors by expressing functional genes of the upstream MEP and MVA pathways, but also the downstream biosynthetic genes and regulators of plastid development, such as GLKs. The data presented here revealed that the insertion of *OsGLK1* in the multigene-stacking expression cassette is an effective strategy to significantly boost carotenoid biosynthesis in rice endosperm. Furthermore, we have noticed that the total carotenoid levels in our GHPC lines are comparable to those of classic *PSY/CrtI* combination

10.3389/fpls.2022.951605

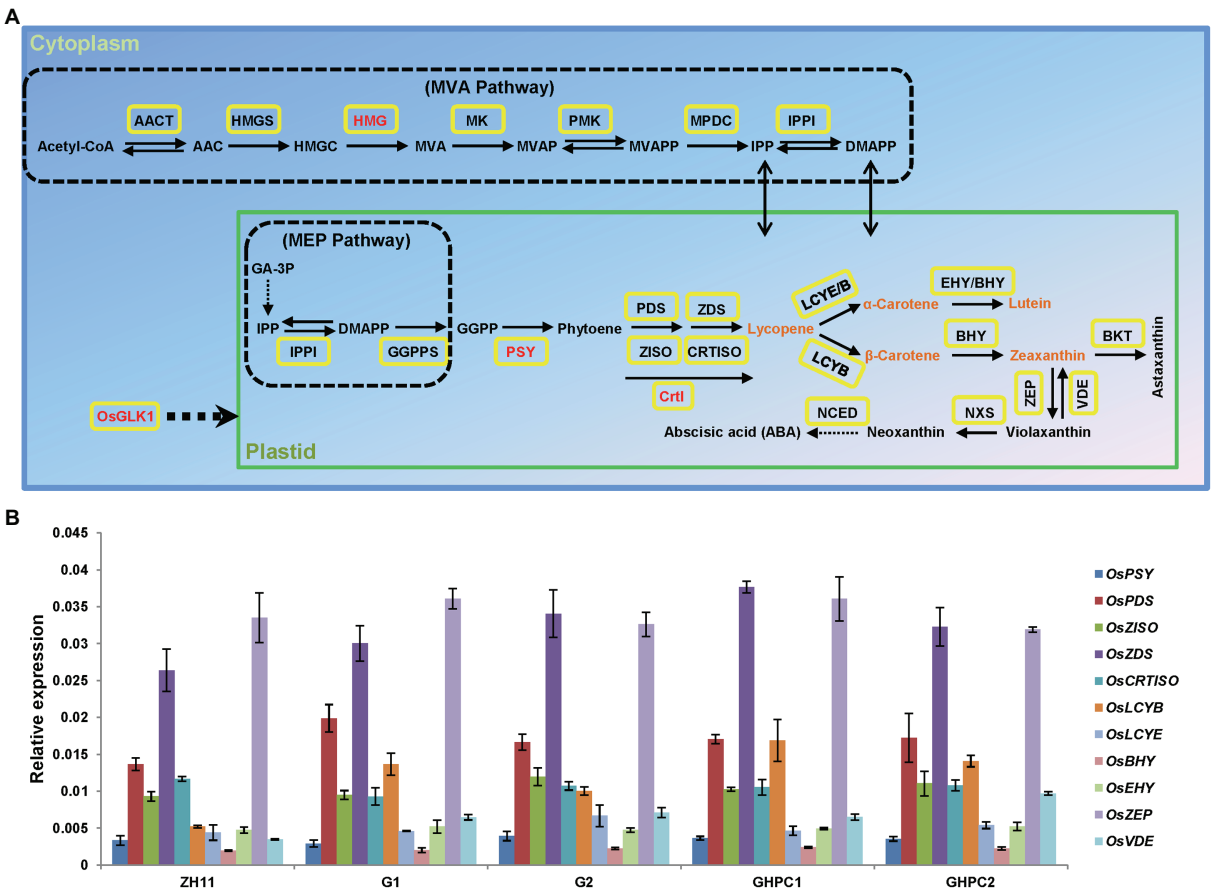


FIGURE 3
Reconstruction of simplified and expanded carotenoid biosynthesis pathway in rice endosperm (Vranová et al., 2013; Nisar et al., 2015) and qRT-PCR analysis of related endogenous carotenogenic genes in ZH11 and transgenic lines. **(A)** Carotenoid biosynthesis pathway combined with MVA pathway in plants. Various conversion steps are indicated with arrows. Yellow boxes represent the genes catalyzing the various reactions. Four transgenes were indicated with red font. AACT, acetoacetyl-CoA thiolase; AAC, acetoacetyl-CoA; HMGS, 3-hydroxy-3-methylglutaryl-CoA synthase; HMGC, 3-hydroxy-3-methylglutaryl-CoA; HMG, 3-hydroxy-3-methylglutaryl-CoA reductase; MVA; mevalonic acid; MK, MVA kinase; MVAP, mevalonate-5-phosphate; PMK, phospho-MVA kinase; MVAPP; mevalonate-5-diphosphate; MPDC, MVA diphosphate decarboxylase; GA-3P, glyceraldehydes 3-phosphate; IPP, isopentenyl diphosphate; IPPI, isopentenyl diphosphate isomerase; DMAPP, dimethylallyl diphosphate; GGPP, geranylgeranyl diphosphate; GGPPS, GGPP synthase; CrtI, bacterial (*Pantoea ananatis*) phytoene desaturase; PSY, phytoene synthase; PDS, phytoene desaturase; ZDS, ξ-carotene desaturase; ZISO, ξ-carotene isomerase; CRTISO, carotenoid isomerase; LCYB, lycopene β-cyclase; LCYE, lycopene ε-cyclase; BHY, β-ring carotene hydroxylase; EHY, ε-ring carotene hydroxylase; BKT, β-carotene ketolase; ZEP, zeaxanthin epoxidase; VDE, violaxanthin de-epoxidase; NXS, neoxanthin synthase; NCED, 9-cis epoxycarotenoid dioxygenase. **(B)** qRT-PCR analysis of endogenous carotenoid biosynthetic genes. RNA samples were prepared from endosperms of developing seeds (9days after pollination) of the homologous T3 lines. The expression levels of the target genes were normalized to the expression of reference gene *OsActin*. Values represent means of three replicates±SD.

(up to 36.7 μg/g; Paine et al., 2005). The carotenoid content in *ZmPSY1/PaCrtI* combination (PC lines) reported by our lab ranged from 7.19 to 8.05 μg/g (Tian et al., 2019). The four genes used in this study were chemically synthesized and the codons were optimized in accordance with the codon usage frequencies of rice. In addition, different plant expression vectors and different transgenic recipient varieties were employed. All these factors combined might have caused a distinction in terms of carotenoid accumulation between our lines and those from Paine et al. (2005). The growing conditions could also contribute to differences in performance.

The expression analysis of endogenous carotenoid biosynthetic genes implicated weak quantitative association between steady-state transcript levels and pigment accumulation. Although most of the carotenoid biosynthetic genes were not differentially expressed in transgenic rice grains compared with controls, the higher plastid numbers and better developed thylakoid presented a source of more carotenoids, which increased the sink capacity for carotenoid accumulation (Lu and Li, 2008). Moreover, the flux of substrate into either branch of xanthophyll formation is controlled by the two *OsLCYs*, converting lycopene into α-carotene and β-carotene. The change in transcript level of *OsLCYB* might contribute to the accumulation of downstream carotenoids. In addition, more lutein than zeaxanthin was found both in G and GHPC lines, which indicated a crucial role of *BHY*. *OsLCYB* and *BHY* could serve as metabolic bottlenecks in the carotenoid biosynthesis

TABLE 2 Grain yield and yield components under normal field conditions in Shanghai and Hainan.

Lines	Shanghai								
	Height (cm)	No. of tillers per plant	Panicle number per plant	Panicle length (cm)	No. of grains per panicle	Filled grains per panicle	Seed-setting rate (%)	1,000-grain weight (g)	Grain yield per plant (g)
ZH11	81.7 ± 2.06	11.50 ± 1.90	10.50 ± 1.58	18.79 ± 1.19	113.80 ± 16.33	93.22 ± 14.70	81.95 ± 5.93	21.08 ± 0.43	19.89 ± 2.99
G1	82.3 ± 0.82	11.83 ± 1.94	11.17 ± 1.60	18.86 ± 0.53	121.42 ± 13.18	113.09 ± 13.09*	93.08 ± 1.34***	20.45 ± 0.26**	24.26 ± 4.07* (21.97%)
G2	81.4 ± 1.51	11.57 ± 1.71	10.71 ± 2.28	18.80 ± 0.59	123.87 ± 7.53	117.11 ± 6.93**	94.56 ± 1.46***	20.53 ± 0.16**	24.37 ± 5.32* (22.52%)
G3	81.5 ± 3.15	11.83 ± 2.23	11.17 ± 2.53	18.47 ± 0.39	118.38 ± 6.84	109.78 ± 5.55*	92.83 ± 3.88**	20.60 ± 0.17*	23.77 ± 6.90* (19.51%)
GHPC1	81.8 ± 4.36	11.67 ± 1.37	10.83 ± 1.47	19.41 ± 0.62	124.84 ± 5.79	114.97 ± 4.87**	92.14 ± 2.61**	20.55 ± 0.14*	23.90 ± 2.76* (20.16%)
GHPC2	81.1 ± 2.97	11.43 ± 1.62	10.57 ± 1.62	19.18 ± 0.58	120.43 ± 5.96	112.45 ± 7.38**	93.34 ± 2.51***	20.43 ± 0.17**	23.39 ± 4.36* (17.6%)
GHPC3	81.8 ± 2.40	12.17 ± 1.33	10.67 ± 1.03	18.95 ± 0.83	121.90 ± 2.38	113.68 ± 2.80**	93.27 ± 1.98***	20.62 ± 0.15*	23.94 ± 2.15* (20.36%)
Lines	**Hainan**								
	Height (cm)	**No. of Tillers per plant**	**Panicle number per plant**	**Panicle length (cm)**	**No. of grains per panicle**	**Filled grains per panicle**	**Seed-setting rate (%)**	**1,000-grain weight (g)**	**Grain yield per plant (g)**
ZH11	82.1 ± 2.33	11.70 ± 1.70	10.70 ± 1.89	19.10 ± 0.88	118.56 ± 13.57	97.69 ± 11.42	82.48 ± 4.81	21.21 ± 0.38	20.01 ± 2.58
G1	82.8 ± 2.14	12.17 ± 1.17	11.67 ± 1.37	19.25 ± 0.45	125.05 ± 6.96	113.05 ± 5.93**	90.49 ± 3.04**	20.57 ± 0.08**	23.96 ± 2.52* (19.74%)
G2	82.4 ± 2.7	11.71 ± 2.14	11.14 ± 2.11	18.97 ± 0.62	119.98 ± 7.08	114.91 ± 6.35**	95.80 ± 1.51***	20.69 ± 0.27**	24.51 ± 4.86* (22.49%)
G3	82.7 ± 1.37	12.50 ± 1.87	11.33 ± 1.97	18.76 ± 0.61	122.94 ± 4.80	116.67 ± 4.26**	94.91 ± 1.23***	20.58 ± 0.17**	24.79 ± 4.48* (23.89%)
GHPC1	83.5 ± 2.94	11.50 ± 1.05	11.17 ± 0.75	19.02 ± 0.36	122.93 ± 5.27	114.76 ± 4.41**	93.40 ± 2.52***	20.73 ± 0.20*	24.43 ± 1.59** (22.09%)
GHPC2	83.1 ± 3.02	12.14 ± 1.77	10.86 ± 1.21	19.17 ± 0.62	119.37 ± 6.29	111.51 ± 6.17*	93.43 ± 2.72***	20.61 ± 0.13**	23.62 ± 3.96* (18.04%)
GHPC3	82.3 ± 1.75	12.33 ± 1.03	10.67 ± 1.03	19.03 ± 0.79	121.99 ± 4.28	115.43 ± 3.57**	94.64 ± 1.26***	20.82 ± 0.21*	24.25 ± 3.53* (21.19%)

Values are mean ± SD ($n = 30$). Asterisks denote Student's t-test significance compared with ZH11 control: *$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$. The percentages in parentheses indicate percentages of grain yield increase compared with ZH11.

pathway, which will help to update strategies for the generation of engineered plants with specific carotenoid profiles or higher levels of carotenoids.

Certain carotenoids, most importantly β-carotene, are cleaved to vitamin A within the body and are referred to as pro-vitamin A (Yeum and Russell, 2002). Vitamin A deficiency, a serious nutritional problem of children all over the developing world, can result in growth retardation in children and seriously impair vision and increase the incidence and severity of infectious diseases (Stephensen, 2001). In Asia, rice is predominantly consumed but lacks pro-vitamin A in the edible part of the grain (endosperm). Biofortification of carotenoids in a staple food such as rice could be of great significance. We all know that an extremely complicated mechanism coordinates carotenoid gene expression, enzyme activities, and plastid differentiation to ensure an appropriate production of carotenoids. In this study, we demonstrate that genetic engineering using *OsGLK1* driven by the endosperm-specific promoter can biofortify the carotenoid biosynthesis to a great extent in the rice endosperm. Our results made a step forward to show the critical influence of plastid identity for carotenoid accumulation. The comparison of metabolic products from GHPC plants and HPC plants from our previous report confirmed that combined expression of the four exogenous genes (*OsGLK1*, *tHMG1*, *ZmPSY1*, and *PaCrtI*) is necessary for attaining higher production of carotenoid in rice endosperm. The GHPC lines reported here has up to 40 μg/g carotenoid, of which nearly 95.7% is α- and β-carotene. This increase in total carotenoid and proportion of α- and β-carotene over the original HPC lines is of great benefit to vitamin A deficiency

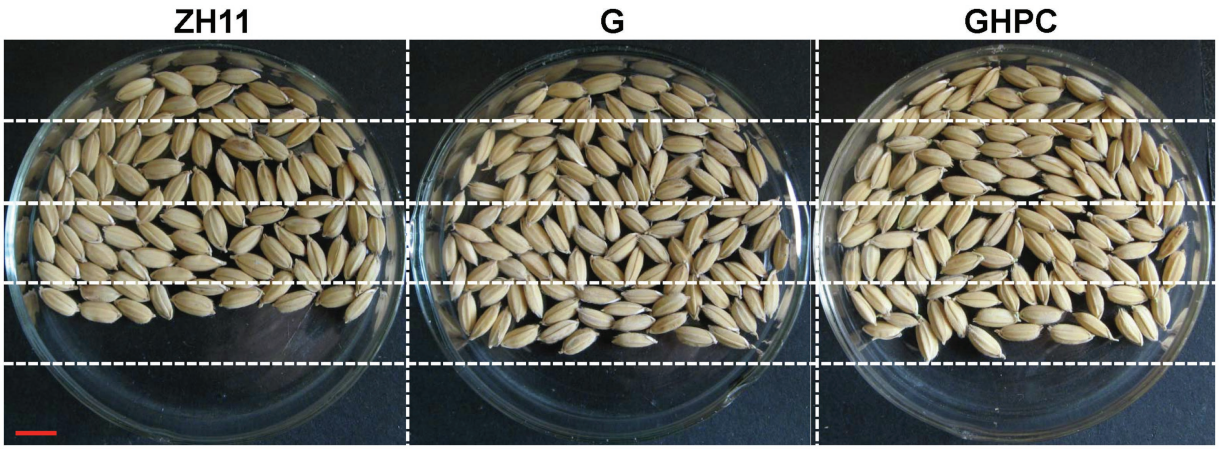


FIGURE 4
Comparison of filled grain number per panicle between ZH11, G and GHPC genotypes (Bar=5mm).

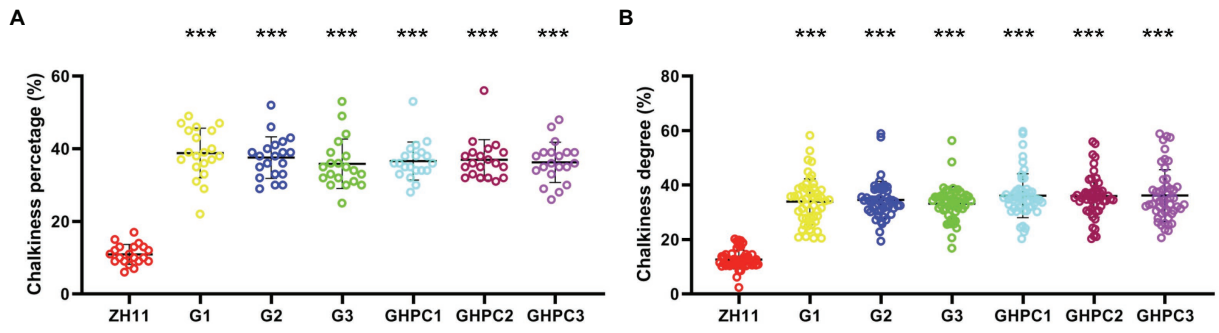


FIGURE 5
Comparison of quality traits between ZH11, G and GHPC lines. **(A)** Percentage of chalkiness (n=20). **(B)** Degree of chalkiness (n=50). Data are given as means±SD. Student's t-test was used to generate the p-values: ***$p < 0.001$.

and related health problems. Moreover, our G and GHPC lines showed increased grain yield and seed-setting rates which were not observed in HPC lines (Tian et al., 2019). It might be safe to conclude that the endosperm-specific overexpression of *OsGLK1* plays a positive role in these two economic traits of rice.

Taken together, the protocol in this study could be used to engineer different biofortified cereals that produce grains enriched with various phytonutrients. This work also facilitates knowledge-based directed biotechnological approaches for complex metabolic engineering in synthetic biology and improvement of quality traits in plants. Overall, our transgenic rice can be used as "cell factories" for producing commercially valuable novel compounds. Our results reveal that enhancing metabolic sink strength for carotenoids in plastids, particularly by promoting plastid formation and development, could hold enormous promise for biofortification of staple crops with increased carotenoid content. Moreover, *OsGLK1* gene might have potential applications for boosting plastid-localized biosynthesis of other nutrients, such as vitamins B2, E and K1, and plastid-dependent enrichment of nutritional elements, such as selenium, in rice grains.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

ZL performed the research and wrote the manuscript. QY, YT, BW, JX, XF, LW, WZ, YD, YW, and ZG critically reviewed and revised the manuscript. JG, RP, and HH analyzed the metabolomics data. All authors contributed to the article and approved the submitted version.

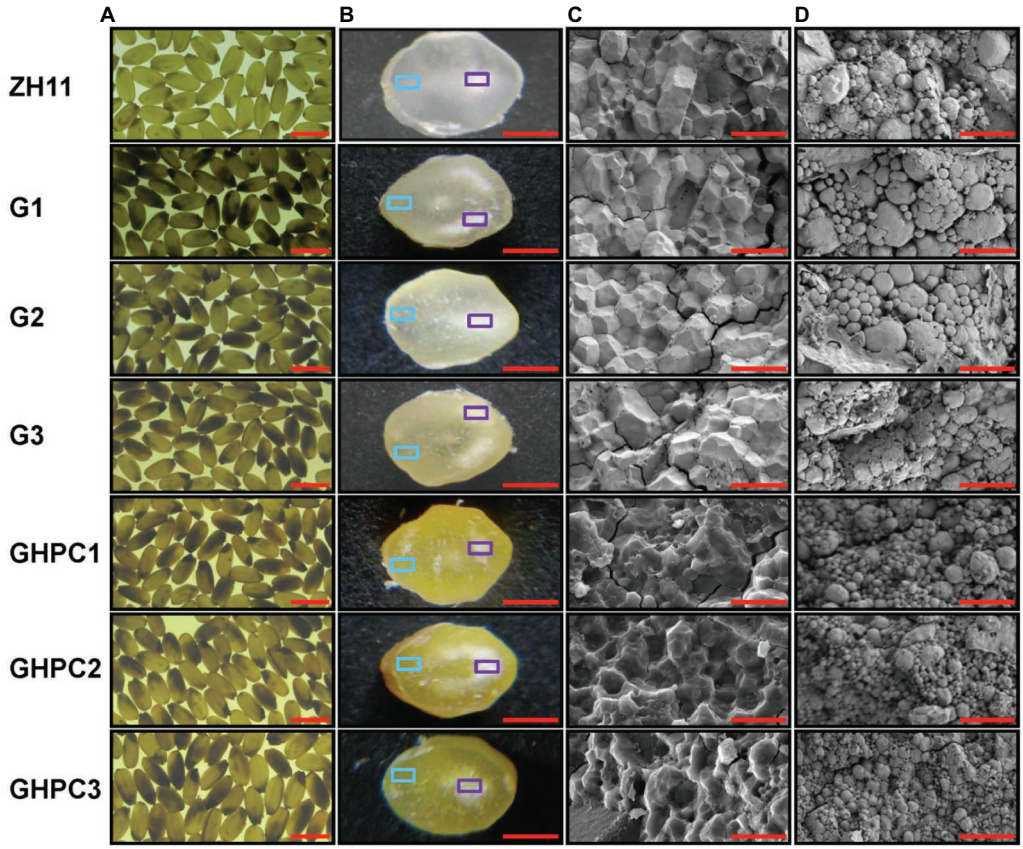


FIGURE 6
Endosperm phenotypes of rice grains of ZH11, G and GHPC plant. **(A)** Mature seeds in white light background. **(B)** Cross sections of mature endosperm. The cross-sections of the sliced grains of the transgenic lines showed that the pigments were evenly distributed in the endosperm of the grains. **(C)** SEM of the mature endosperm, with the cross sections indicated by a blue square in **(B)**. **(D)** SEM of the mature endosperm, with the cross sections indicated by a purple square in **(B)**. SEM, scanning electron microscopy. Scale bars: 5mm **(A)**; 1mm **(B)**; 20μm **(C,D)**.

Funding

This work was supported by Innovation Team Project of Shanghai Academy of Agricultural Sciences [2022–005] and Shanghai Academic Technology Research Leader (19XD1432300). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments

We thank Lei Yang (Biotree biotech Co., Ltd., Shanghai, China) for technical assistance with statistical analysis of metabolomics data.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.951605/full#supplementary-material

SUPPLEMENTARY FIGURE S1

HPLC analysis of carotenoids in mature rice endosperms from ZH11, G and GHPC lines. Labeled peaks: A, astaxanthin; 1, lutein; 2, zeaxanthin; 3, lycopene, 4, α-carotene; 5, β-carotene. AU, absorbance unit.

SUPPLEMENTARY FIGURE S2

Hierarchical cluster analysis (HCA) of 89 differentially produced metabolites in mature ZH11 and G endosperms. Metabolite content is presented as median-centered averages with three biological replicates each. Red and blue colors indicate high and low content, respectively.

SUPPLEMENTARY FIGURE S3

Expression levels of the three transgenes in endosperms of two HPC lines. RNA samples were prepared from endosperms of developing seeds (9days after pollination) of the homologous T3 lines. The transcripts were normalized to the expression of reference gene *OsActin*. Values represent means of three replicates±SD.

SUPPLEMENTARY FIGURE S4

Transmission electron micrographs of rice endosperm. **(A)** The endosperm cells (15 DAP) of HPC genotype are shown. Scale bar=1μm. **(B)** The endosperm cells (45 DAP) of HPC genotype are shown. Scale bar=2μm. PL, plastids; SG, starch grain; OD, oil droplet; CW, cell walls.

SUPPLEMENTARY FIGURE S5

Transmission electron micrographs of rice seed coats. **(A)** Phenotypes of rice seed coats from ZH11, G, and GHPC lines. Scale bar=1mm. **(B)** The cells of rice seed coats are shown. Scale bar=5μm. **(C)** Chloroplasts boxed in (B) are shown by higher magnification. Scale bar=1μm. The seed coats were sampled from 15 DAP rice grains of indicated genotypes.

References

Auldridge, M. E., Block, A., Vogel, J. T., Dabney-Smith, C., Mila, I., Bouzayen, M., et al. (2006). Characterization of three members of the Arabidopsis carotenoid cleavage dioxygenase family demonstrates the divergent roles of this multifunctional enzyme family. *Plant J.* 45, 982–993. doi: 10.1111/j.1365-313X.2006.02666.x

Bai, C., Capell, T., Berman, J., Medina, V., Sandmann, G., Christou, P., et al. (2016). Bottlenecks in carotenoid biosynthesis and accumulation in rice endosperm are influenced by the precursor–product balance. *Plant Biotechnol. J.* 14, 195–205. doi: 10.1111/pbi.12373

Bai, C., Twyman, R. M., Farré, G., Sanahuja, G., Christou, P., Capell, T., et al. (2011). A golden era—pro-vitamin A enhancement in diverse crops. *In Vitro Cell. Dev. Biol. Plant* 47, 205–221. doi: 10.1007/s11627-011-9363-6

Bouis, H. E., and Saltzman, A. (2017). Improving nutrition through biofortification: a review of evidence from HarvestPlus, 2003 through 2016. *Glob. Food Sec.* 12, 49–58. doi: 10.1016/j.gfs.2017.01.009

Campbell, R., Ducreux, L. J. M., Morris, W. L., Morris, J. A., Suttle, J. C., Ramsay, G., et al. (2010). The metabolic and developmental roles of carotenoid cleavage dioxygenase 4 from potato. *Plant Physiol.* 154, 656–664. doi: 10.1104/pp.110.158733

Cazzonelli, C. I., and Pogson, B. J. (2010). Source to sink: regulation of carotenoid biosynthesis in plants. *Trends Plant Sci.* 15, 266–274. doi: 10.1016/j.tplants.2010.02.003

Chaudhary, N., Nijhawan, A., Khurana, J. P., and Khurana, P. (2009). Carotenoid biosynthesis genes in rice: structural analysis, genome-wide expression profiling and phylogenetic analysis. *Mol. Gen. Genomics.* 283:13. doi: 10.1007/s00438-009-0495-x

Chen, M., Ji, M., Wen, B., Liu, L., Li, S., Chen, X., et al. (2016). GOLDEN 2-LIKE transcription factors of plants. *Front. Plant Sci.* 7:1509. doi: 10.3389/fpls.2016.01509

De Steur, H., Blancquaert, D., Strobbe, S., Lambert, W., Gellynck, X., and Van Der Straeten, D. (2015). Status and market potential of transgenic biofortified crops. *Nat. Biotechnol.* 33, 25–29. doi: 10.1038/nbt.3110

Enfissi, E. M. A., Fraser, P. D., Lois, L.-M., Boronat, A., Schuch, W., and Bramley, P. M. (2005). Metabolic engineering of the mevalonate and non-mevalonate isopentenyl diphosphate-forming pathways for the production of health-promoting isoprenoids in tomato. *Plant Biotechnol. J.* 3, 17–27. doi: 10.1111/j.1467-7652.2004.00091.x

Fang, X., Mao, Y., and Chen, X. (2018). Engineering purple rice for human health. *Sci. China Life Sci.* 61, 365–367. doi: 10.1007/s11427-017-9157-7

Farré, G., Sanahuja, G., Naqvi, S., Bai, C., Capell, T., Zhu, C., et al. (2010). Travel advice on the road to carotenoids in plants. *Plant Sci.* 179, 28–48. doi: 10.1016/j.plantsci.2010.03.009

Fitter, D. W., Martin, D. J., Copley, M. J., Scotland, R. W., and Langdale, J. A. (2002). GLK gene pairs regulate chloroplast development in diverse plant species. *Plant J.* 31, 713–727. doi: 10.1046/j.1365-313X.2002.01390.x

Fraser, P. D., and Bramley, P. M. (2004). The biosynthesis and nutritional uses of carotenoids. *Prog. Lipid Res.* 43, 228–265. doi: 10.1016/j.plipres.2003.10.002

Giuliano, G. (2014). Plant carotenoids: genomics meets multi-gene engineering. *Curr. Opin. Plant Biol.* 19, 111–117. doi: 10.1016/j.pbi.2014.05.006

Hall, L. N., Rossini, L., Cribb, L., and Langdale, J. A. (1998). GOLDEN 2: a novel transcriptional regulator of cellular differentiation in the maize leaf. *Plant Cell* 10, 925–936. doi: 10.1105/tpc.10.6.925

Hemmerlin, A., Harwood, J. L., and Bach, T. J. (2012). A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Prog. Lipid Res.* 51, 95–148. doi: 10.1016/j.plipres.2011.12.001

Hirschberg, J. (2001). Carotenoid biosynthesis in flowering plants. *Curr. Opin. Plant Biol.* 4, 210–218. doi: 10.1016/S1369-5266(00)00163-1

Hirschi, K. D. (2009). Nutrient biofortification of food crops. *Annu. Rev. Nutr.* 29, 401–421. doi: 10.1146/annurev-nutr-080508-141143

Huang, J.-C., Zhong, Y.-J., Liu, J., Sandmann, G., and Chen, F. (2013). Metabolic engineering of tomato for high-yield production of astaxanthin. *Metab. Eng.* 17, 59–67. doi: 10.1016/j.ymben.2013.02.005

Imai, K. K., Ohashi, Y., Tsuge, T., Yoshizumi, T., Matsui, M., Oka, A., et al. (2006). The a-type Cyclin CYCA2;3 is a key regulator of Ploidy levels in *Arabidopsis* Endoreduplication. *Plant Cell* 18, 382–396. doi: 10.1105/tpc.105.037309

Langdale, J. A., and Kidner, C. A. (1994). Bundle sheath defective, a mutation that disrupts cellular differentiation in maize leaves. *Development* 120, 673–681. doi: 10.1242/dev.120.3.673

Le Gall, G., Dupont, M. S., Mellon, F. A., Davis, A. L., Collins, G. J., Verhoeyen, M. E., et al. (2003). Characterization and content of flavonoid glycosides in genetically modified tomato (*Lycopersicon esculentum*) fruits. *J. Agric. Food Chem.* 51, 2438–2446. doi: 10.1021/jf025995e

Li, Z., Fu, X., Tian, Y., Xu, J., Gao, J., Wang, B., et al. (2019). Overexpression of a trypanothione synthetase gene from Trypanosoma cruzi, TcTrys, confers enhanced tolerance to multiple abiotic stresses in rice. *Gene* 710, 279–290. doi: 10.1016/j.gene.2019.06.018

Liu, J., Liu, Y., Jia, M., Kang, X., Wang, S., Sun, H., et al. (2021). Association of enriched metabolites profile with the corresponding volatile characteristics induced by rice yellowing process. *Food Chem.* 349:129173. doi: 10.1016/j.foodchem.2021.129173

Lopez-Juez, E., and Pyke, K. A. (2005). Plastids unleashed: their development and their integration in plant development. *Int. J. Dev. Biol.* 49, 557–577. doi: 10.1387/ijdb.051997el

Lu, S., and Li, L. (2008). Carotenoid metabolism: biosynthesis, regulation, and beyond. *J. Integr. Plant Biol.* 50, 778–785. doi: 10.1111/j.1744-7909.2008.00708.x

Luo, D., Xu, H., Liu, Z., Guo, J., Li, H., Chen, L., et al. (2013). A detrimental mitochondrial-nuclear interaction causes cytoplasmic male sterility in rice. *Nat. Genet.* 45, 573–577. doi: 10.1038/ng.2570

Mangel, N., Fudge, J. B., Li, K.-T., Wu, T.-Y., Tohge, T., Fernie, A. R., et al. (2019). Enhancement of vitamin B6 levels in rice expressing Arabidopsis vitamin B6 biosynthesis de novo genes. *Plant J.* 99, 1047–1065. doi: 10.1111/tpj.14379

Mitrophanov, A. Y., and Groisman, E. A. (2008). Positive feedback in cellular control systems. *Bioessays* 30, 542–555. doi: 10.1002/bies.20769

Nguyen, C. V., Vrebalov, J. T., Gapper, N. E., Zheng, Y., Zhong, S., Fei, Z., et al. (2014). Tomato *GOLDEN2-LIKE* transcription factors reveal molecular gradients that function during fruit development and ripening. *Plant Cell* 26, 585–601. doi: 10.1105/tpc.113.118794

Nisar, N., Li, L., Lu, S., Khin, N. C., and Pogson, B. J. (2015). Carotenoid metabolism in plants. *Mol. Plant* 8, 68–82. doi: 10.1016/j.molp.2014.12.007

Paine, J. A., Shipton, C. A., Chaggar, S., Howells, R. M., Kennedy, M. J., Vernon, G., et al. (2005). Improving the nutritional value of Golden Rice through increased pro-vitamin A content. *Nat. Biotechnol.* 23, 482–487. doi: 10.1038/nbt1082

Paul, J.-Y., Khanna, H., Kleidon, J., Hoang, P., Geijskes, J., Daniells, J., et al. (2017). Golden bananas in the field: elevated fruit pro-vitamin A from the expression of a single banana transgene. *Plant Biotechnol. J.* 15, 520–532. doi: 10.1111/pbi.12650

Peng, R., Fu, X., Tian, Y., Zhao, W., Zhu, B., Xu, J., et al. (2014). Metabolic engineering of Arabidopsis for remediation of different polycyclic aromatic hydrocarbons using a hybrid bacterial dioxygenase complex. *Metab. Eng.* 26, 100–110. doi: 10.1016/j.ymben.2014.09.005

Rodríguez-Concepción, M. (2010). Supply of precursors for carotenoid biosynthesis in plants. *Arch. Biochem. Biophys.* 504, 118–122. doi: 10.1016/j.abb.2010.06.016

Rossini, L., Cribb, L., Martin, D. J., and Langdale, J. A. (2001). The maize *Golden2* gene defines a novel class of transcriptional regulators in plants. *Plant Cell* 13, 1231–1244. doi: 10.1105/tpc.13.5.1231

Ruiz-Sola, M., and Rodríguez-Concepción, M. (2012). Carotenoid biosynthesis in Arabidopsis: a colorful pathway. *Arabidopsis Book* 10:e0158. doi: 10.1199/tab.0158

Schaub, P., Al-Babili, S., Drake, R., and Beyer, P. (2005). Why is Golden Rice Golden (yellow) instead of red? *Plant Physiol.* 138, 441–450. doi: 10.1104/pp.104.057927

Sharma, S., Khare, P., Kumar, A., Chunduri, V., Kumar, A., Kapoor, P., et al. (2020). Anthocyanin-biofortified colored wheat prevents high fat diet–induced alterations in mice: nutrigenomics studies. *Mol. Nutr. Food Res.* 64, 1900999. doi: 10.1002/mnfr.201900999

Shewmaker, C. K., Sheehy, J. A., Daley, M., Colburn, S., and Ke, D. Y. (1999). Seed-specific overexpression of phytoene synthase: increase in carotenoids and other metabolic effects. *Plant J.* 20, 401–412. doi: 10.1046/j.1365-313x.1999.00611.x

Stephensen, C. B. (2001). Vitamin A, infection, and immune function. *Annu. Rev. Nutr.* 21, 167–192. doi: 10.1146/annurev.nutr.21.1.167

Sun, T., Yuan, H., Cao, H., Yazdani, M., Tadmor, Y., and Li, L. (2018). Carotenoid metabolism in plants: the role of plastids. *Mol. Plant* 11, 58–74. doi: 10.1016/j.molp.2017.09.010

Tian, Y., Wang, B., Peng, R., Xu, J., Li, T., Fu, X., et al. (2019). Enhancing carotenoid biosynthesis in rice endosperm by metabolic engineering. *Plant Biotechnol. J.* 17, 849–851. doi: 10.1111/pbi.13059

Underwood, B. A., and Arthur, P. (1996). The contribution of vitamin A to public health. *FASEB J.* 10, 1040–1048. doi: 10.1096/fasebj.10.9.8801165

Vranová, E., Coman, D., and Gruissem, W. (2013). Network analysis of the MVA and MEP pathways for Isoprenoid synthesis. *Annu. Rev. Plant Biol.* 64, 665–700. doi: 10.1146/annurev-arplant-050312-120116

Washida, H., Wu, C.-Y., Suzuki, A., Yamanouchi, U., Akihama, T., Harada, K., et al. (1999). Identification of cis-regulatory elements required for endosperm expression of the rice storage protein glutelin gene GluB-1. *Plant Mol. Biol.* 40, 1–12. doi: 10.1023/A:1026459229671

Wu, Z., Luo, Y., Bao, J., Luo, Y., and Yu, Z. (2020). Additives affect the distribution of metabolic profile, microbial communities and antibiotic resistance genes in high-moisture sweet corn kernel silage. *Bioresour. Technol.* 315:123821. doi: 10.1016/j.biortech.2020.123821

Xiong, A.-S., Yao, Q.-H., Peng, R.-H., Duan, H., Li, X., Fan, H.-Q., et al. (2006). PCR-based accurate synthesis of long DNA sequences. *Nat. Protoc.* 1, 791–797. doi: 10.1038/nprot.2006.103

Yasumura, Y., Moylan, E. C., and Langdale, J. A. (2005). A conserved transcription factor mediates nuclear control of organelle biogenesis in anciently diverged land plants. *Plant Cell* 17, 1894–1907. doi: 10.1105/tpc.105.033191

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P., et al. (2000). Engineering the Provitamin A (β-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287, 303–305. doi: 10.1126/science.287.5451.303

Yeum, K. J., and Russell, R. M. (2002). Carotenoid bioavailability and bioconversion. *Annu. Rev. Nutr.* 22, 483–504. doi: 10.1146/annurev.nutr.22.010402.102834

Yuan, H., Zhang, J., Nageswaran, D., and Li, L. (2015). Carotenoid metabolism and regulation in horticultural crops. *Hortic. Res.* 2, 15036. doi: 10.1038/hortres.2015.36

Zhai, S., Xia, X., and He, Z. (2016). Carotenoids in staple cereals: metabolism, regulation, and genetic manipulation. *Front. Plant Sci.* 7:1197. doi: 10.3389/fpls.2016.01197

Zhu, C., Naqvi, S., Breitenbach, J., Sandmann, G., Christou, P., and Capell, T. (2008). Combinatorial genetic transformation generates a library of metabolic phenotypes for the carotenoid pathway in maize. *Proc. Natl. Acad. Sci. U. S. A.* 105, 18232–18237. doi: 10.1073/pnas.0809737105

Zhu, C., Naqvi, S., Gomez-Galera, S., Pelacho, A. M., Capell, T., and Christou, P. (2007). Transgenic strategies for the nutritional enhancement of plants. *Trends Plant Sci.* 12, 548–555. doi: 10.1016/j.tplants.2007.09.007

Zhu, Q., Zeng, D., Yu, S., Cui, C., Li, J., Li, H., et al. (2018). From Golden Rice to aSTARice: bioengineering Astaxanthin biosynthesis in Rice endosperm. *Mol. Plant* 11, 1440–1448. doi: 10.1016/j.molp.2018.09.007

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 30 September 2022
DOI 10.3389/fpls.2022.968780

OPEN ACCESS

EDITED BY
Zhi-Yan (Rock) Du,
University of Hawaii at Manoa,
United States

REVIEWED BY
Min Shi,
Zhejiang Chinese Medical University, China
Saeid Kadkhodaei,
Agricultural Biotechnology Research Institute of Iran, Iran

*CORRESPONDENCE
Qiang Zhuge
qzhuge@njfu.edu.cn
Liming Yang
yangliming@njfu.edu.cn
Xiaohong Zhou
20190024@zafu.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Plant Systems and Synthetic Biology, a section of the journal Frontiers in Plant Science

RECEIVED 14 June 2022
ACCEPTED 22 August 2022
PUBLISHED 30 September 2022

CITATION
Movahedi A, Wei H, Pucker B, Ghaderi-Zefrehei M, Rasouli F, Kiani-Pouya A, Jiang T, Zhuge Q, Yang L and Zhou X (2022) Isoprenoid biosynthesis regulation in poplars by methylerythritol phosphate and mevalonic acid pathways.
Front. Plant Sci. 13:968780.
doi: 10.3389/fpls.2022.968780

Isoprenoid biosynthesis regulation in poplars by methylerythritol phosphate and mevalonic acid pathways

Ali Movahedi[1†], Hui Wei[2†], Boas Pucker[3†], Mostafa Ghaderi-Zefrehei[4], Fatemeh Rasouli[5,6], Ali Kiani-Pouya[5,6], Tingbo Jiang[7], Qiang Zhuge[1*], Liming Yang[1*] and Xiaohong Zhou[8*]

[1]Key Laboratory of Forest Genetics and Biotechnology, Ministry of Education, Co-Innovation Center for Sustainable Forestry in Southern China, College of Biology and the Environment, Nanjing Forestry University, Nanjing, China, [2]Key Laboratory of Landscape Plant Genetics and Breeding, School of Life Sciences, Nantong University, Nantong, China, [3]Institute of Plant Biology and BRICS, TU Braunschweig, Braunschweig, Germany, [4]Department of Animal Science, Faculty of Agriculture, Yasouj University, Yasouj, Iran, [5]State Key Laboratory of Molecular Plant Genetics, Shanghai Center for Plant Stress Biology, Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai, China, [6]Tasmanian Institute of Agriculture, College of Science and Engineering, University of Tasmania, Hobart, TAS, Australia, [7]State Key Laboratory of Tree Genetics and Breeding, Northeast Forestry University, Harbin, China, [8]State Key Laboratory of Subtropical Silviculture, Zhejiang A&F University, Hangzhou, Zhejiang, China

It is critical to develop plant isoprenoid production when dealing with human-demanded industries such as flavoring, aroma, pigment, pharmaceuticals, and biomass used for biofuels. The methylerythritol phosphate (MEP) and mevalonic acid (MVA) plant pathways contribute to the dynamic production of isoprenoid compounds. Still, the cross-talk between MVA and MEP in isoprenoid biosynthesis is not quite recognized. Regarding the rate-limiting steps in the MEP pathway through catalyzing 1-deoxy-D-xylulose5-phosphate synthase and 1-deoxy-D-xylulose5-phosphate reductoisomerase (DXR) and also the rate-limiting step in the MVA pathway through catalyzing 3-hydroxy-3-methylglutaryl-CoA reductase (HMGR), the characterization and function of *HMGR* from *Populus trichocarpa* (*PtHMGR*) were analyzed. The results indicated that *PtHMGR* overexpressors (OEs) displayed various MEP and MVA-related gene expressions compared to NT poplars. The overexpression of *PtDXR* upregulated MEP-related genes and downregulated MVA-related genes. The overexpression of *PtDXR* and *PtHMGR* affected the isoprenoid production involved in both MVA and MEP pathways. Here, results illustrated that the *PtHMGR* and *PtDXR* play significant roles in regulating MEP and MVA-related genes and derived isoprenoids. This study clarifies cross-talk between MVA and MEP pathways. It demonstrates the key functions of *HMGR* and *DXR* in this cross-talk, which significantly contribute to regulate isoprenoid biosynthesis in poplars.

KEYWORDS

methylerythritol phosphate pathway, mevalonic acid pathway, HMGR, DXR, isoprenoid biosynthesis, poplar

Introduction

Isoprenoids (terpenoids) present many functional and structural properties over 50,000 distinct molecules in living organisms (Thulasiram et al., 2007). Isoprenoids play vital roles in plant growth and development, as well as in membrane fluidity, photosynthesis, and respiration. They are joining with plant-pathogen and allelopathic interactions to preserve plants from pathogens and herbivores, and they are also created to draw pollinators and seed-dispersing animals. The importance of isoprenoids has been proved in rubber products, drugs, flavors, fragrances, agrochemicals, nutraceuticals, disinfectants, and pigments (Bohlmann and Keeling, 2008). A wide range of isoprenoids is involved in photosynthesis in plants, including electron transfer, quenching of excited chlorophyll triplets, light-harvesting, and energy conversion (Malkin and Niyogi, 2000). Tetrapyrrole ring derived from heme pathway with an appended isoprenoid-derived phytol chain aim chlorophylls in all reaction centers and antenna complexes to absorb light energy and transfer electrons to the reaction centers. The light-harvesting system is maintained by the other isoprenoids, which are linear or partly cyclized carotenes and their oxygenated derivatives, xanthophylls. These carotenes and xanthophylls reduce excess excitation energy through the process of light-harvesting. However, these isoprenoids are attractive in flowers and fruits (Rodriguez-Concepcion, 2010). In contrast to vertebrates synthesizing cholesterol, higher plants synthesize a complex mix of sterol lipids called phytosterols (Boutte and Grebe, 2009).

Plant isoprenoids include gibberellins (GAs), carotene, lycopene, cytokinins (CKs), strigolactones (GRs), and brassinosteroids (BRs), which are produced through methylerythritol phosphate (MEP) and mevalonic acid (MVA) pathways (van Schie et al., 2006; Xie et al., 2008; Henry et al., 2015). These pathways are involved in plant growth, development, and response to environmental stresses (Bouvier et al., 2005; Kirby and Keasling, 2009). The isopentenyl diphosphate isomerase (IDI) catalyzes the conversion of the isopentenyl diphosphate (IPP) into dimethylallyl diphosphate (DMAPP), leading to provide the basic precursors for all isoprenoid productions (Hemmerlin et al., 2012; Lu et al., 2012; Zhang et al., 2019). The IPP and DMAPP are essential for regulating MEP and MVA pathway interactions (Huchelmann et al., 2014; Liao et al., 2016). The MVA pathway reactions appear in the cytoplasm, endoplasmic reticulum (ER), and peroxisomes (Cowan et al., 1997; Roberts, 2007), producing sesquiterpenoids and sterols. The 3-hydroxy-3-methylglutaryl-coenzyme A synthase (HMGR), a rate-limiting enzyme in the MVA pathway, catalyzes 3-hydroxy-3-methylglutary-CoA (HMG-CoA) to form MVA (Cowan et al., 1997; Roberts, 2007).

On the other hand, the MEP pathway reactions appear in the chloroplast, producing carotenoids, GAs, and diterpenoids. The 1-deoxy-D-xylulose5-phosphate synthase (DXS) and 1-deoxy-D-xylulose5-phosphate reductoisomerase (DXR) are rate-limiting enzymes in the MEP pathway that catalyze the conversion of glyceraldehyde3-phosphate (G-3-P) and pyruvate (Cordoba et al., 2009; Wang et al., 2012; Yamaguchi et al., 2018; Perreca et al., 2020). Isoprenoids like phytoalexin and volatile oils play essential roles in plant growth, development, and disease resistance (Hain et al., 1993; Ren et al., 2008). Besides an extensive range of natural functions in plants, terpenoids also consider the potential for biomedical applications. Paclitaxel is one of the most effective chemotherapy agents for cancer treatment, and artemisinin is an anti-malarial drug (Kong and Tan, 2015; Kim et al., 2016).

Previous metabolic engineering studies have proposed strategies to improve the production of specific plant metabolites (Ghirardo et al., 2014; Opitz et al., 2014). In addition, it has been proved that *HMGS* overexpression upregulated carotenoid and phytosterol in tomatoes (Liao et al., 2018). The HMGR has been considered as a critical factor in the metabolic engineering of terpenoids (Aharoni et al., 2005; Dueber et al., 2009), and its overexpression in ginseng (*PgHMGR1*) increased ginsenosides content, which is a necessary pharmaceutically active component (Kim et al., 2014). Overexpression of *Salvia miltiorrhiza HMGR* significantly improves diterpenoid tanshinone contents (Dai et al., 2011; Kai et al., 2011). The first generation of transgenic tomatoes exhibited a higher phytosterol content because of the overexpression of *HMGR1* from *Arabidopsis thaliana* (*AtHMGR1*; Enfissi et al., 2005).

Further studies about the MEP pathway revealed that the dissemination of *DXR* transcript resulted in decreased pigmentation amount, while its upregulation caused the accumulation of the MEP-derived isoprenoids in plastid (Carretero-Paulet et al., 2006). The overexpression of *DXR from A. thaliana* caused to accumulation of the diterpene anthiolimine in *Salvia sclarea* hairy roots (Vaccaro et al., 2014). In another study, the overexpression of *DXR* from peppermint resulted in approximately 50% increased monoterpene compositions (Mahmoud and Croteau, 2001). Further studies on MVA and MEP pathways exhibited their cross-talk in metabolic intermediates through plastid membranes (Laule et al., 2003; Liao et al., 2006).

Poplars provide the raw materials for industrial and agricultural production as an economic and energy species. Its fast growth characteristics and advanced resources in artificial afforestation play a vital role in the global ecosystem (Devappa et al., 2015). This study investigates the biosynthesis of isoprenoids in poplar. The increased expression of *PtHMGR* in *Populus trichocarpa* increased the transcript levels of genes associated with MVA and MEP pathways. This work further exhibits that the overexpression of *PtDXR* in *P. trichocarpa* causes downregulating of genes related to MVA and, conversely, upregulating of genes associated with MEP. The overexpression of *PtDXR* also affects GA_3, trans-zeatin-riboside (TZR), isopentenyl adenosine (IPA),

Abbreviations: MEP, Methylerythritol phosphate; MVA, Mevalonic acid; DXR, 1-deoxy-D-xylulose5-phosphate reductoisomerase; HGMR, 3-hydroxy-3-methylglutaryl-CoA reductase; GAs, Gibberellins; CKs, Cytokinins; GRs, Strigolactones; BRs, Brassinosteroids; IDI, Isopentenyl diphosphate isomerase; IPP, Isopentenyl diphosphate; DMAPP, Dimethylallyl diphosphate; HMG-CoA, 3-hydroxy-3-methylglutary-CoA; DXS, 1-deoxy-D-xylulose5-phosphate synthase; G-3-P, Glyceraldehyde3-phosphate.

castasterone (CS), and 6-deoxocastasterone (DCS) amounts. Taken together, *PtHMGR* and *PtDXR* genes play critical regulatory points in MEP and MVA pathways through isoprenoid productions.

Materials and methods

Plant materials and growth conditions

The "Nanlin 895" (*Populus deltoides*×*Populus euramericana* "Nanlin895") plants were cultured in half-strength Murashige and Skoog (1/2 MS) medium (pH 5.8) under conditions of 24°C and 74% humidity (Movahedi et al., 2015). Subsequently, NT (Transformed WT with empty vector) and transgenic poplars were cultured in 1/2 MS under 16/8 h light/dark at 24°C for 1 month (Movahedi et al., 2018).

PtHMGR gene amplification and vector construction

The total RNA was extracted from *P. trichocarpa* leaves and processed with PrimeScript™ RT Master Mix, a kind of reverse transcriptase (TaKaRa, Japan). Forward and reverse primers (Supplementary Table 1: *PtHMGR*-F and *PtHMGR*-R) were designed, and the open reading frame (ORF) of *PtHMGR* was amplified *via* PCR. The total volume of 50 µl, including 2 µl primers (10 µM), 2.0 µl cDNA (200 ng), 5.0 µl 10 × PCR buffer (Mg^{2+}), 4 µl dNTPs (2.5 mM), 0.5 µl rTaq polymerase (5 U; TaKaRa, Japan) was then used for the following PCR reactions: 95°C for 7 min, 35 cycles of 95°C for 1 min, 58°C for 1 min, 72°C for 1.5 min, and 72°C for 10 min. Subsequently, the product of the *PtHMGR* gene was ligated into the pEASY-T3 vector (TransGen Biotech, China) based on blue-white spot screening, and the *PtHMGR* gene was inserted into the vector pGWB9 (Song et al., 2016) using Gateway technology (Invitrogen, United States).

Phylogenetic analyses

The National Center for Biotechnology Information database (NCBI) has been applied to download *HMGR* from *P. trichocarpa* (*PtHMGR*: Potri.004G208500; XM_006384809) and from other 34 species to align and analyze. The alignment has been performed using Geneious Prime Ver. 2022 software (Biomatters development team) to construct a phylogenetic tree with 1,000 bootstrap replicates.

Plant transformation

PtHMGR transgenic generation and confirmation

The CDS of *PtHMGR* was cloned in pGWB9 vector, and the *Agrobacterium tumefaciens* var. EHA105, including recombinant pGWB9-*PtHMGR,* was used to transform "Nanlin 895" poplar leaves and petioles (Movahedi et al., 2014; Zhang et al., 2017). Poplar buds were screened on differentiation MS medium supplemented with 30 µg/ml Kanamycin (Kan). Resistant buds were planted in the bud elongation MS medium containing 20 µg/ml Kan and transplanted into 1/2 MS medium, including 10 µg/ml Kan to generate resistant poplar trees (Movahedi et al., 2014). Genomic DNA has been extracted from putative transformants of one-month-old leaves grown on medium-supplemented kanamycin using TianGen kits (TianGen BioTech, China). The quality of the extracted genomic DNA (250–350 ng/µl) was determined by a BioDrop spectrophotometer (United Kingdom). PCR was then conducted using designed primers (Supplementary Table 1: CaMV35S as the forward and *PtHMGR*-R as the reverse), Easy Taq polymerase (TransGene Biotech), and 50 ng of extracted genomic DNA as a template to amplify about 2,000 bp. In addition, total RNA was extracted from leaves to produce cDNA, as mentioned above. These cDNAs were then applied to real-time quantitative PCR (qPCR; Supplementary Table 1: *PtHMGR* forward and reverse) to evaluate the *PtHMGR* expression in *PtHMGR* overexpressors (*OEs*) compared with NT poplars. Three independent technical repeats were performed to analyze *PtHMGR* expression.

PtDXR transgenic supply and confirmation

PtDXR transformant seedlings were provided from Nanjing Xiaozhuang University through the key laboratory of quality and safety of agricultural products, which has been sequenced previously and confirmed (Xu et al., 2019), *followed by culturing under the same conditions as PtHMGR poplars. PtDXR transformants were further confirmed through PCR confirmation for T-DNA insertion.*

Non-transformant poplars

NT poplars were provided by transforming empty vectors into WT poplars for both PtHMGR and PtDXR transformants, followed by culturing with no medium-supplemented kanamycin to screen.

Phenotypic properties evaluation

The 75-day-old *PtHMGR*-OEs, *PtDXR*-OEs, and NT poplars were selected to evaluate the phenotypic changes such as stem lengths (mm) and diameters (mm). *PtHMGR*-OE3 and-OE7 (Two plants) and *PtDXR*-OE1 and-OE3 (Two plants) were screened 45 times during 75 days to evaluate and compare with NT in the same conditions (Started on day 5th and continued randomly until the 75th day). All data were analyzed by GraphPad Prism 9, using ANOVA one way (Supplementary Table 2).

Analyses of expression of MVA and MEP-related genes with qRT-PCR

Three-month-old *PtDXR-OEs, PtHMGR-OEs*, and NT (grown on soil) leaves were used to extract the total RNA. The

qPCR was performed to identify MVA and MEP-related gene expressions in *PtDXR-OE* and *PtHMGR-OE*s compared to NT. We used the StepOne Plus Real-time PCR System (Applied Biosystems, United States) and SYBR Green Master Mix (Roche, Germany) to carry out qRT-PCR with the *PtActin* gene (XM-006370951) serving as the housekeeping gene standard (Zhang et al., 2013). The following conditions were used for qPCR reactions: pre-denaturation at 95°C for 10 min, 40 cycles of denaturation at 95°C for 15 s, and a chain extension at 60°C for 1 min. Three independent biological replicates were used with three technical repeats (Supplementary Table 1; *PtHMGR* and *DXR* forward and reverse).

Quantitative detection of endogenous hormone contents

The AB Qtrap6500 mass spectrometer was used in triple four-stage rod-ion hydrazine mode. The HPLC-MS/MS method was used to quantitatively analyze GA_3, TZR, IPA, DCS, and CS hormones. Samples were separated using Agilent 1290 HPLC with electrospray ionization as the ion source and scanned in a multi-channel detection mode. Three-month-old leaves were used for hormone extraction. Liquid nitrogen was used to grind fresh plant samples of 0.5 g, and then a mixture of 10 ml isopropanol and hydrochloric acid was added to shake for 30 min at 4°C. Next, 20 ml of dichloromethane was added to the solution, which was shaken for 30 min at 4°C. Subsequently, the collected solution was centrifuged at 13,000× g for 20 min at 4°C, and the lower organic phase was retained. Then, the organic phase was dried under nitrogen and dissolved in 400 µl methanol containing 0.1% formic acid. Finally, the collected solution was filtered through a 0.22 µm membrane and detected using HPLC-MS/MS.

The standard solutions with variable concentrations of 0.1, 0.2, 0.5, 2, 5, 20, 50, and 200 ng/ml were prepared with methanol (0.1% formic acid) as the solvent (Supplementary Figures 5–9). During plotting of the standard curve, non-linear points can be excluded. In the liquid phase, a Poroshell 120 SB-C18 column (2.1 × 150, 2.7 m) was used at a column temperature of 30°C. The mobile phase included A: B = (methanol/0.1% formic acid): (water/0.1% formic acid). Elution gradient: 0–1 min, A = 20%; 1–9 min, A = 80%; 9–10 min, A = 80%; 10–10.1 min, A = 20%; 10.1–15 min, A = 20%. The other requirements used in this experiment included 2 µl injection volume, 15 psi air curtain gas, 4,500 v spray voltage, 65 psi atomization pressure, 70 psi auxiliary pressure, and 400°C atomization temperature.

Quantitative determination of carotenoids and lycopene contents

A sample of 1 g of poplar leaves was ground and dissolved in a mixture of 10 ml of acetone-petroleum ether (1:1). Subsequently, the supernatant was filtered and collected repeatedly. The collected solution was transferred to a liquid separation funnel and layered statically. Then, the upper organic phase was extracted through a funnel containing anhydrous sodium sulfate, and the lower layer was extracted again in petroleum ether. The collected solution was evaporated to dry with rotation at 35°C, dissolved into 1 ml dichloromethane, and filtered. Finally, the solution was filtered through a 0.45-µm filter membrane and analyzed with HPLC. In liquid phase conditions, the Waters Symmetry Shield RP18 reversed-phase chromatographic column (4.6 × 250 mm × 5 µm) was used in this study with a column temperature of 30°C. The mobile phase was a mixture of methanol, acetonitrile, and dichloromethane (methanol: acetonitrile: dichloromethane = 20: 75:5). Elution gradient: 0–30 min, mobile phase = 100%, flow rate = 1.0 ml/min. The injection volume was 10 µl. In addition, the standard curves of α-carotenoid, β-carotenoid, and lycopene were formulated as described above.

Results

PCR and real-time PCR confirmed the transformation of *PtHMGR*

The high similarity of achieved alignment, including lots of conserved amino acids (aa), similar domains, HMG-CoA-binding motifs (EMPVGYVQIP and TTEGCLVA), and NADPH-binding motifs (DAMGMNMV and VGTVGGGT; Ma et al., 2012), confirmed detected PtHMGR protein analytically (Supplementary Figure 1). A phylogenetic tree based on the HMGRs from various species supported the PtHMGR candidate identification (Supplementary Figure 2). The tblastn was then applied to reveal 2,614 bp *PtHMGR* on Chr04:21681480…21,684,242 with a 1,785 bp CDS. The PCR amplified 1855 bp of the *PtHMGR* from synthesized cDNA of *P. trichocarpa* confirmed the putative transgenic lines (Supplementary Figure 3A), exhibiting amplicons in PCR identification compared with NT poplar (Supplementary Figure 3B). The expression of *PtHMGR* through *PtHMGR-OEs* was significantly higher than in NT poplars (Supplementary Figure 3C), indicating successful overexpression of *PtHMGR* in poplar.

The overexpression of *PtHMGR* and *PtDXR* regulate MVA and MEP-related genes

MVA-related genes such as acetoacetyl CoA thiolase (*AACT*), mevalonate kinase (*MVK*), mevalonate5-diphosphate decarboxylase (*MVD*), *HMGR*, and farnesyl diphosphate synthase *(FPS)*, except *HMGS*, were significantly upregulated in *PtHMGR-OEs* (Figure 1A; Supplementary Figure 4A). In contrast, *PtDXR* overexpression caused upregulating of the *FPS* expression and downregulating of the *AACT*, *HMGS*, *HMGR*,

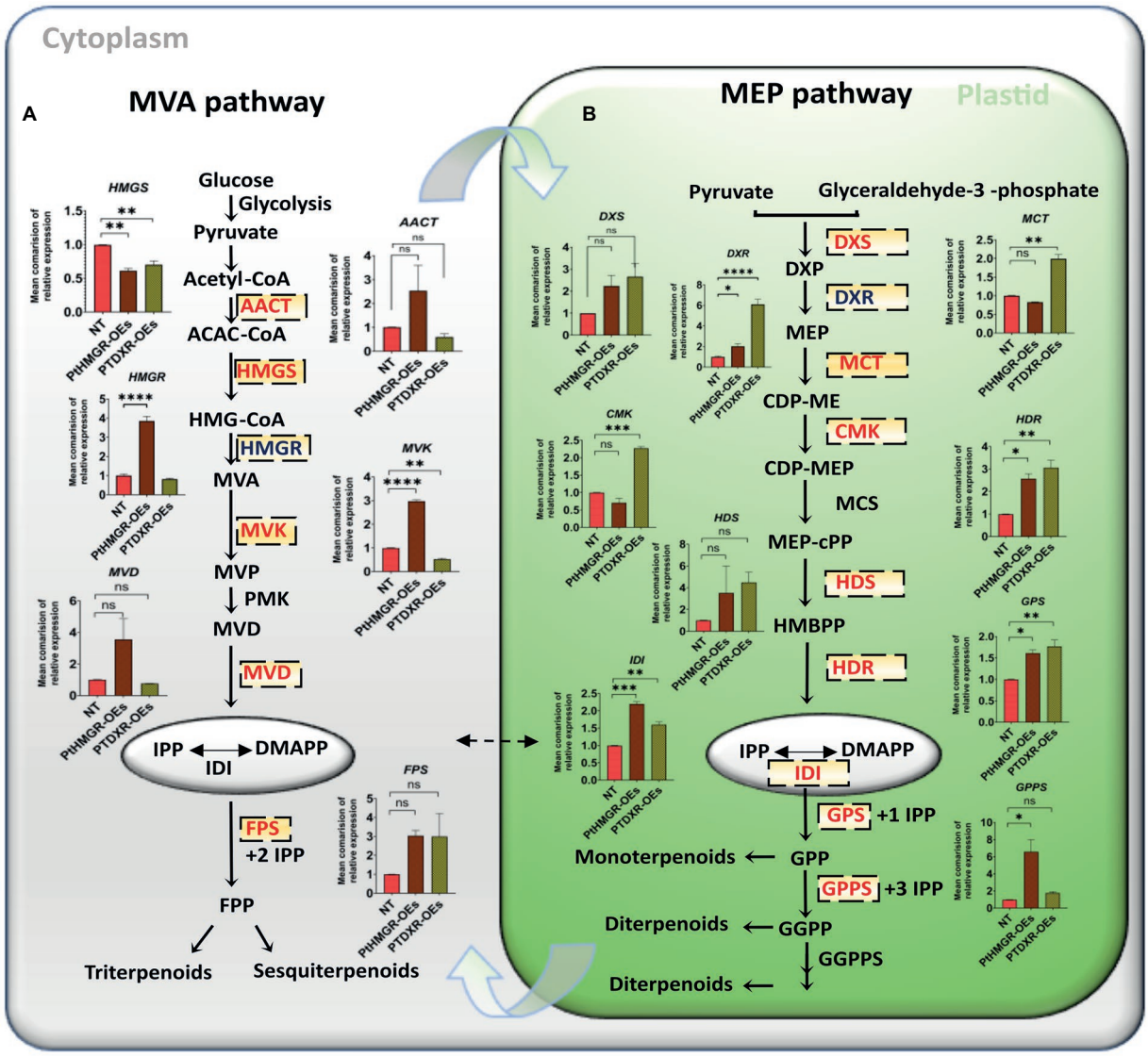


FIGURE 1
MVA and MEP-related gene expression analyses in PtHMGR-and PtDXR-OEs poplars. **(A)** Mean comparison of relative expressions of MVA related genes affected by PtHMGR-and PtDXR-OEs. **(B)** Mean comparison of relative expressions of MEP related genes affected by PtHMGR-and PtDXR-OEs; PtActin was used as an internal reference in all repeats; "ns" means not significant, *p <0.05, **p <0.01, ***p <0.001, ****p <0.0001; Three independent replications were performed in each experiment.

MVD, and *MVK* expressions (Figure 1A; Supplementary Figure 4B). On the other hand, *PtHMGR*-OEs caused significantly upregulate the MEP-related genes such as *DXS*, *DXR*, 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate synthase (*HDS*), 1-hydroxy-2-methyl-2-(E)-butenyl-4-diphosphate reductase (*HDR*), *IDI*, geranyl pyrophosphate synthase *(GPS)*, and geranyl diphosphate synthase (*GPPS;* Figure 1B; Supplementary Figure 4C). In contrast, *PtHMGR-OEs* caused downregulating 2-C-methyl-d-erythritol4-phosphate cytidylyltransferase (*MCT*) and 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (*CMK*). Moreover, *PtDXR* overexpression caused upregulating all MEP-related genes (Figure 1B; Supplementary Figure 4D).

PtHMGR positively impacts MVA and MEP-derived carotenoid contents

β-carotene is a carotenoid synthesis that has been broadly used in the industrial composition of pharmaceuticals and as food colorants, animal supplies additives, and nutraceuticals. In addition, lycopene is a precursor of β-carotene, referring to C40 terpenoids, and is broadly found in various plants, particularly vegetables and fruits. It has been shown that MVA and MEP pathways directly influence the biosynthesis production of lycopene (Wei et al., 2018; Kim et al., 2019). While one study showed that β-carotene and lutein are synthesized using intermediates from the MEP pathway (Wille et al., 2004), the other study revealed that both MVA and

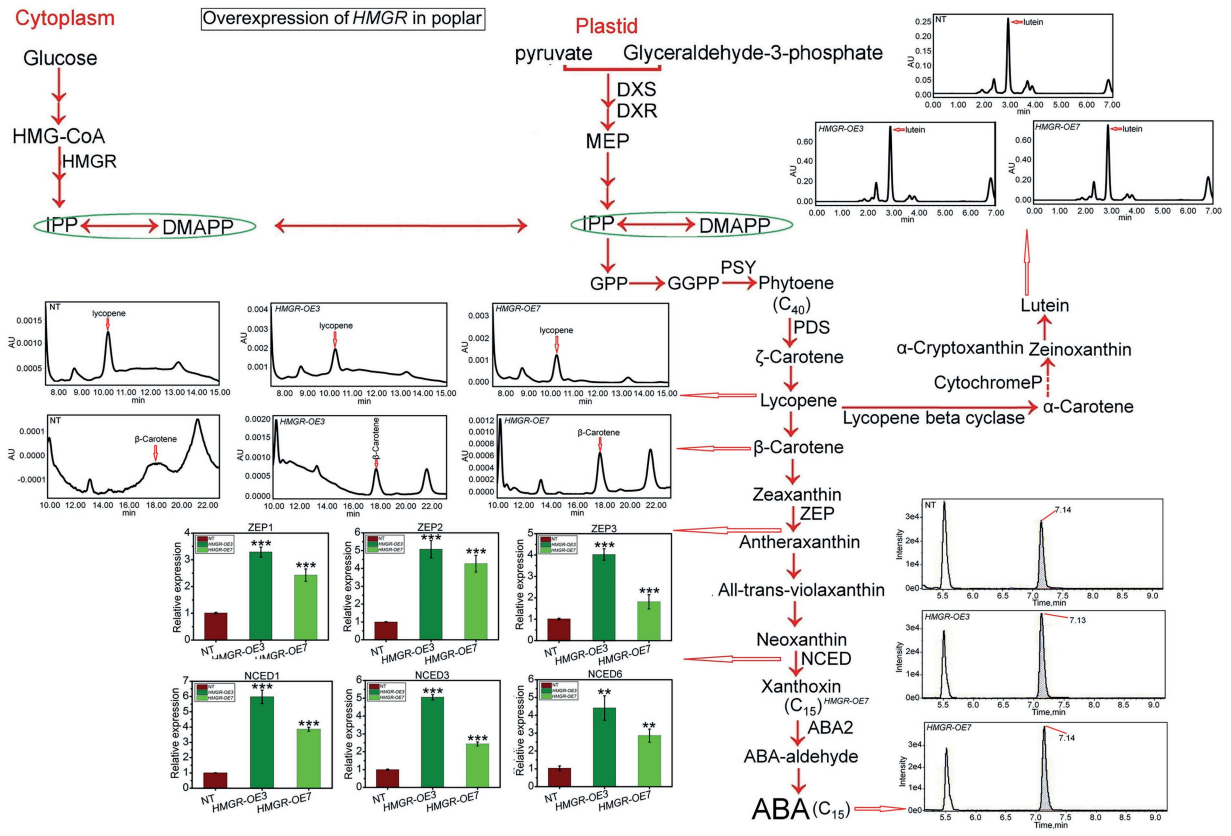


FIGURE 2
Evaluation of the effect of the overexpression of *PtHMGR* on isoprenoid production and ABA. The qRT-PCR analyses of the expression levels of ABA synthesis-related genes and HPLC-MS/MS analyses of the contents of ABA, β-carotene, lycopene, and lutein in *PtHMGR*-OEs. The expression levels of the ABA synthesis-related *ZEP* and *NCED* genes family are shown. Bars represent mean±SD (n =3); Stars reveal significant differences, $^{**}p$ <0.01, $^{***}p$ <0.001; Experiments were performed in triplicates.

MEP pathways produce isoprenoids such as β-carotene and lutein (Opitz et al., 2014). HPLC-MS/MS has been applied to analyze the MVA and MEP derivatives to quantity. Our analyses revealed a slight increase in the average lycopene content through *PtHMGR*-OE3 and-OE7 transformants compared to NT (0.0013 AU) with 0.0016 AU (Figure 2). The overexpression of *HMGR* caused significantly enhanced the synthesis of β-carotene with an average of 0.00065 AU compared with NT by-0.00002 AU (Figure 2). Results also exhibited a significant increase in lutein content through PtHMGR-OE3 and-OE7 with an average of 0.7 AU compared with NT with 0.25 AU (Figure 2). Moreover, the abscisic acid (ABA) content has been influenced by overexpression of *HMGR* with a considerable increase with an average of 3.7e4 compared with NT with 2.7e4 (Figure 2). The ABA-related gene expressions were also calculated, and results revealed that the overexpression of *HMGR* significantly upregulated the zeaxanthin epoxidase (*ZEP*) *1*, *−2*, and −3 genes with averages of ~2.8, ~4.6, and ~2.9, respectively in comparing with NT poplars (Figure 2). These results also indicated meaningful upregulation through 9-cis-epoxycarotenoid dioxygenase (*NCED*) 1, *−3*, and −6 genes with averages of ~4.16, ~3.79, and ~3.4 compared with NT with an average of ~1 (Figure 2).

In addition, the total level of GA_3, a downstream product of MEP, increased to 0.2–0.35 ng/g in *PtHMGR*-OE lines, compared to 0.08–0.1 ng/g in NT poplars (Figure 3A). Also, the IPA content in *PtHMGR*-OEs (0.32–0.41 ng/g) is dominantly higher than that in NT (0.54–0.66 ng/g; Figure 3B). Among MVA-derived isoprenoids, the results of HPLC-MS/MS exhibited significant enhancements in the TZR content through *PtHMGR*-OEs with an average of 0.56 ng/g than NT with an average of 0.3 ng/g (Figure 3C). Surprisingly, the expression of *HMGR* caused a 3-fold increase in DCS content with an average of 5 ng/g compared with NT of about 1.5 ng/g (Figure 3D). However, *PtHMGR*-OEs negatively affected the CS content with a significant decrease compared to NT poplars (Figure 3E).

PtDXR influence on MVA and MEP-derived carotenoid contents positively

PtDXR-OEs exhibited a significant increase in lycopene, β-carotene, and lutein contents affected by overexpression of

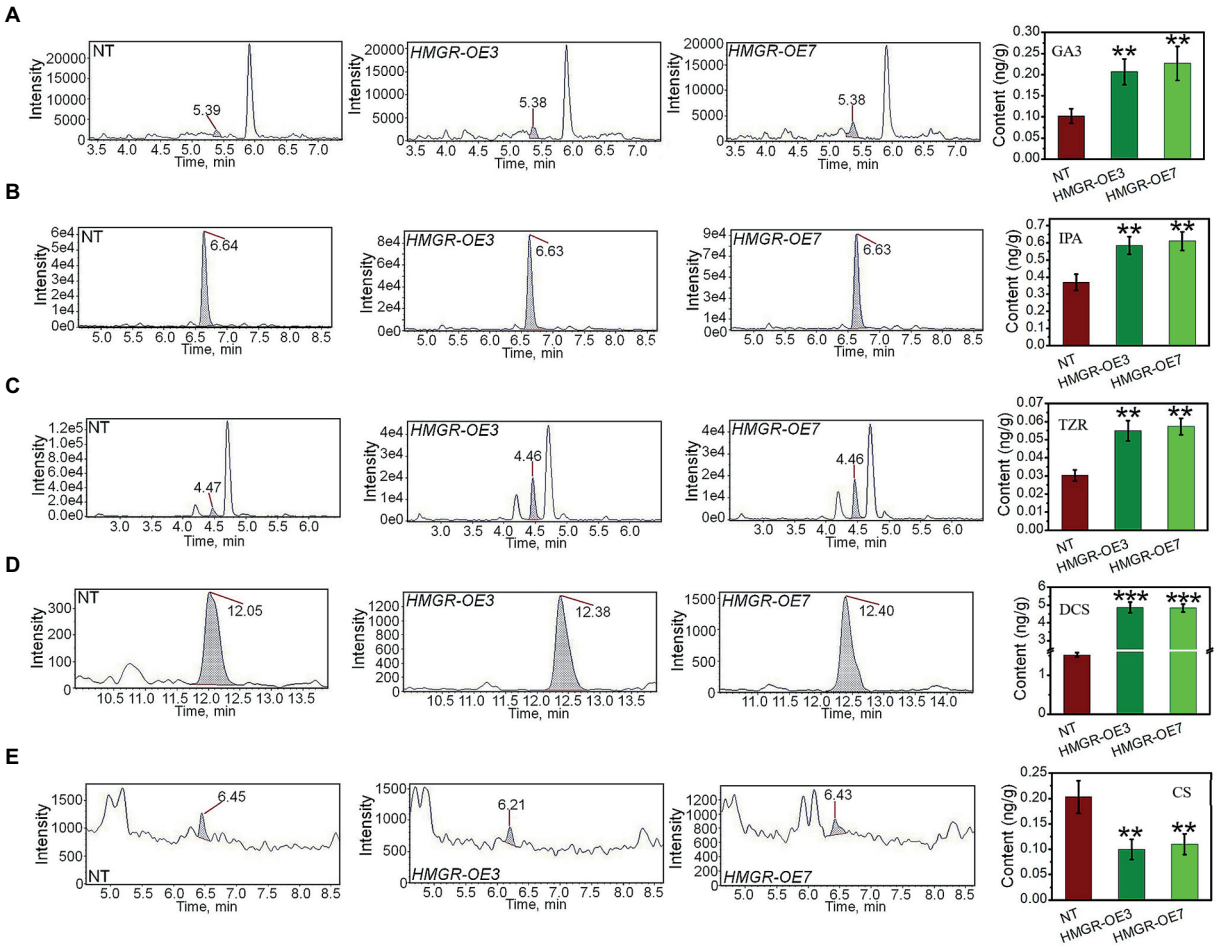


FIGURE 3
Analyzing the effect of the overexpression of *PtHMGR* on MVA and MEP-derivatives. HPLC-MS/MS analyses of the contents of GA_3 **(A)**, IPA **(B)**, TZR **(C)**, DCS **(D)**, and CS **(E)** in *PtHMGR-OEs* compared to NT poplars. Bars represent mean±SD ($n = 3$); Stars reveal significant differences, $^{**}p < 0.01$, $^{***}p < 0.001$; Experiments were performed in triplicates.

PtDXR with the averages of 0.00095, 0.00037, and 0.5 AU compared to NT poplars (Figure 4). Moreover, the ABA content also has been affected by the overexpression of *PtDXR* and considerably increased more than NT poplars (Figure 4). The expression of the ABA-related gene also has been influenced by the overexpression of *PtDXR* and upregulated *NCED* genes, except for *NCED1*, which has been downregulated (Figure 4). These analyses further exhibited the positive effect of the overexpression of *PtDXR* on a significant increase in the expression of *ZEP1*, *2*, and *3* (Figure 4).

The further analysis of the production of GA_3 exhibited a significant increase affected by the overexpression of *PtDXR* with an average of 0.27 ng/g compared to NT poplars (Figure 5A). Also, the overexpression of *PtDXR* caused to increase in IPA production significantly (Figure 5B). A considerable increase in TZR content also was observed, influenced by the overexpression of *PtDXR* with an average of 0.3 ng/g compared to NT with only 0.04 ng/g (Figure 5C). In addition, the overexpression of *PtDXR* caused to increase in the DCS production significantly, but adversely the CS was meaningfully decreased (Figures 5D,E).

Pleiotropic analyzes

To improve the effect of overexpression *of PtHMGR* and *PtDXR* on phenotypic changes, the stem length and stem diameter have been evaluated. Regarding the results, the overexpression of *PtHMGR* and *PtDXR* revealed positive effects on increasing the isoprenoid contents. More increase in cytokinins TZR (~0.290 ng/g) and IPA (~0.940 ng/g) affected by the overexpression of *PtDXR* in comparing with TZR (~0.054 ng/g) and IPA (~0.580 ng/g) affected by the overexpression of *PtHMGR*, resulting in significantly more developments in the stem length through *PtDXR*-OEs compared with *PtHMGR*-OEs (Figures 6A,B). Slightly more stem diameters through *PtDXR*-OEs compared with *PtHMGR*-OEs, may be resulted from increasing the expression of ABA-related genes

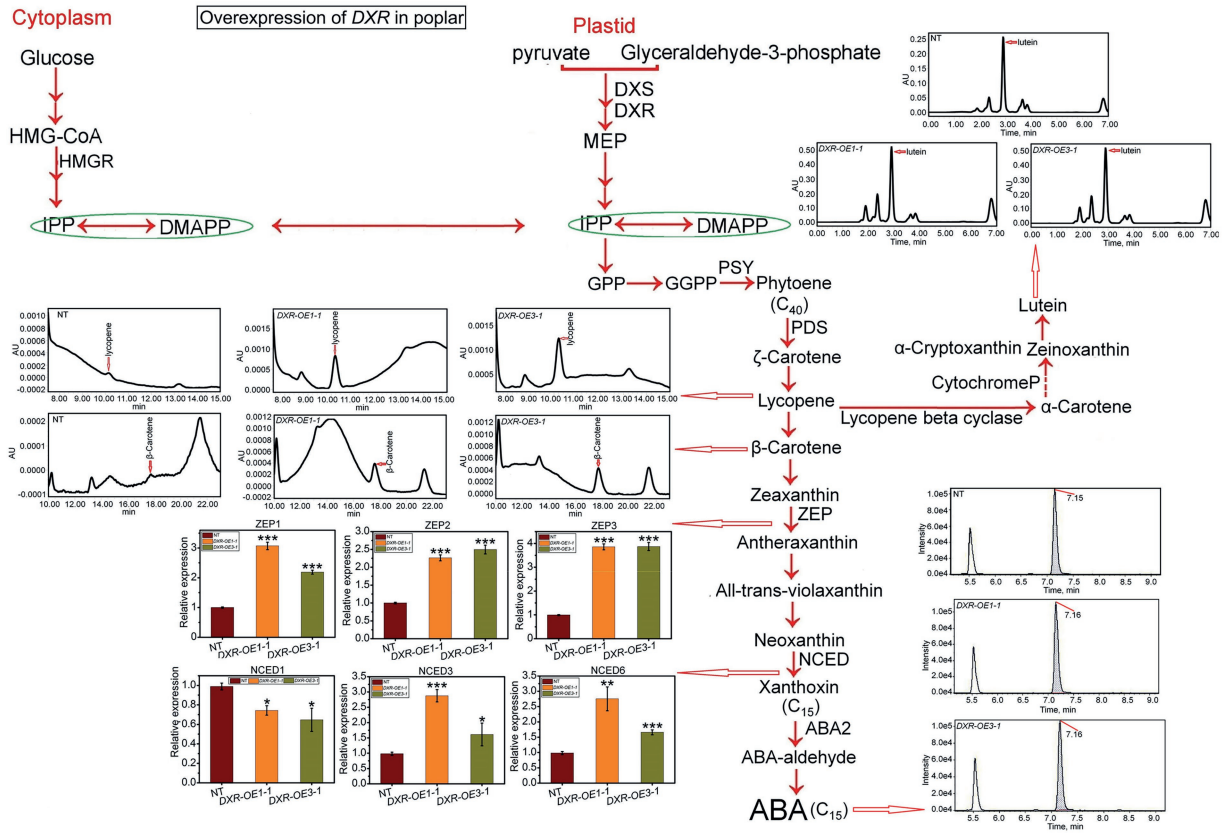


FIGURE 4
Assessment of the impact of *PtDXR* overexpression on isoprenoid and ABA biosynthesis. The qRT-PCR analyses of the expression levels of ABA synthesis-related genes and HPLC-MS/MS analyses of the contents of ABA, β-carotene, lycopene, and lutein in *PtDXR*-OEs. The expression levels of the ABA synthesis-related *ZEP* and *NCED* genes family are indicated. Bars represent mean±SD (n=3); Stars reveal significant differences, *p <0.05, **p <0.01, ***p <0.001; Experiments were performed in triplicates.

ZEP (~3.40) and *NCED* (~3.93) in *PtHMGR-OEs* compared with *ZEP* (~2.70) and *NCED* (~1.57) in *PtDXR-OEs* (Figures 6C–E).

Discussion

The HMGR and DXR revealed crucial roles in isoprenoid biosynthesis

Several studies report that the activity of HMGR is regulated by isoprenoid outcomes when stigmasterol and cholesterol reduce the HMGR activity by 35% (Russell and Davidson, 1982). The other study revealed a downregulation of the activity of HMGR in pea by approximately 40% among those treated with ABA, while zeatin and gibberellin improved the activity of HMGR (Russell and Davidson, 1982). Moreover, the MEP pathway is the primary precursor for required plastid isoprenoids (Maurey et al., 1986; Wright et al., 2014). It has been demonstrated that volatile compounds produced by the MEP pathway are involved in protecting plants against biotic and abiotic stresses (Gershenzon and Dudareva, 2007). Additionally, promising metabolite compounds have been developed in the mint plant by modifying the expression of *DXR* (Mahmoud and Croteau, 2002).

Overexpression of *DXR* in Arabidopsis results in accumulations of isoprenoids such as tocopherols, carotenoids, and chlorophylls (Carretero-Paulet et al., 2006). Overexpression of *DXR* has also been proven to improve diterpene content in transgenic bacteria (Morrone et al., 2010). Biotic tolerance, which is important in providing pharmaceutical terpenoids by expanding the number of enzymes included in biosynthetic pathways (Kang et al., 2009; Lu et al., 2016), caused improved *DXR* expression followed by triptophenolide content in *Tripterygium wilfordii* cell culture suspension (Tong et al., 2015).

Overexpression of *PtHMGR* and *PtDXR* regulate the transcription of MEP and MVA-related genes

It has been shown that the overexpression of *BjHMGS1* (MVA-related gene) affected the expressions of MEP-related

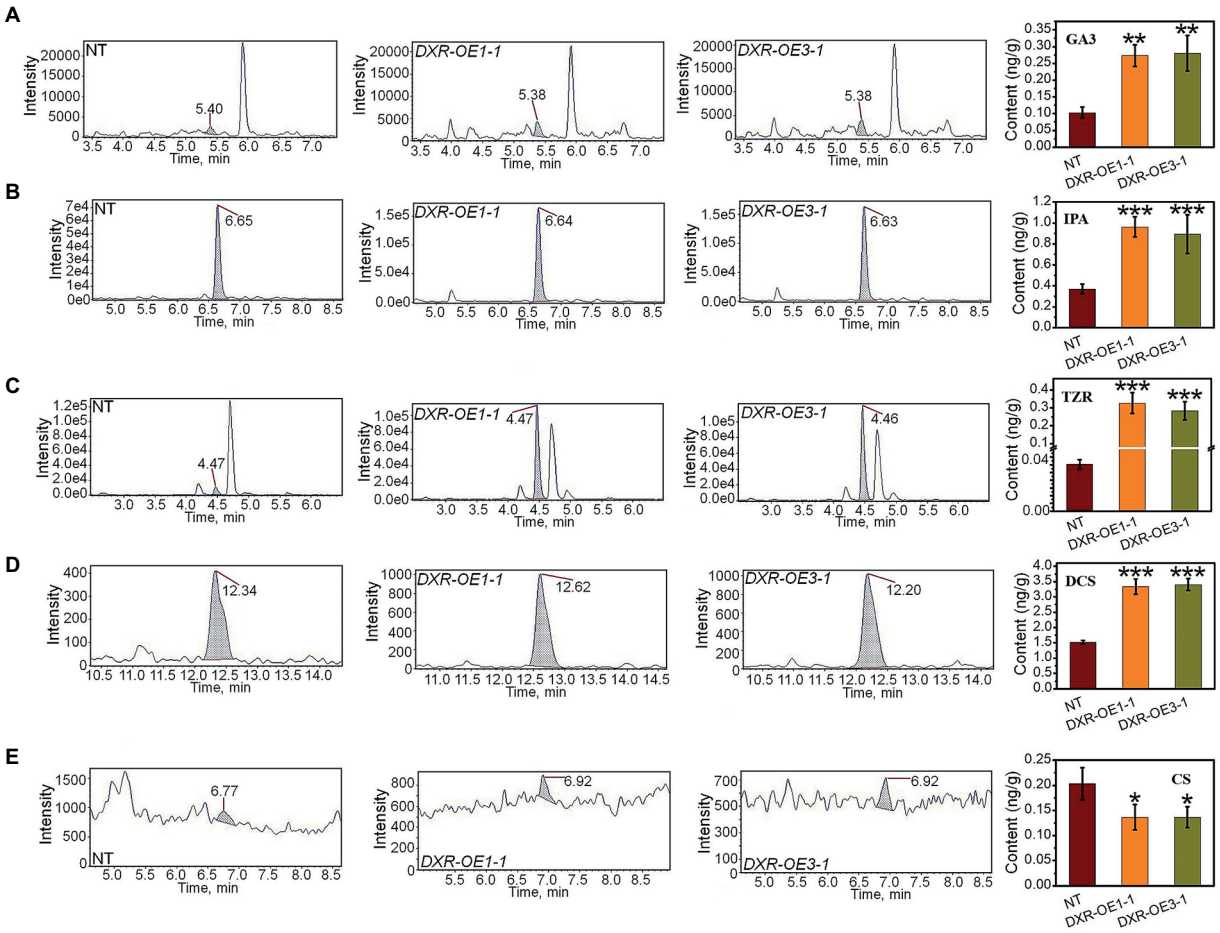


FIGURE 5
Analyzing the effect of the overexpression of *PtDXR* on MVA and MEP-derivatives. HPLC-MS/MS analyses of the contents of GA_3 **(A)**, IPA **(B)**, TZR **(C)**, DCS **(D)**, and CS **(E)** in *PtDXR-OEs* compared to NT poplars.

genes and slightly increased the transcript levels of *DXS* and *DXR* in transgenic tomatoes (Liao et al., 2018). In this study, *PtHMGR-OEs* exhibited the upregulation of the MEP-related genes *DXS*, *DXR*, *HDS*, and *HDR* and the downregulation of *MCT* and *CMK*. Like *BjHMGS1* overexpression in tomatoes that demonstrated a significantly upregulated of the MEP-related genes *GPS* and *GPPS* (Liao et al., 2018), we also indicated that the overexpression of *PtHMGR* enhanced the *GPS*, and *GPPS* expressions, resulting in stimulating the crosstalk between IPP and DMAPP, which increased the biosynthesis of plastidial C15 and C20 isoprenoid precursors. Moreover, it has been shown that the overexpression *of HMGR* in *Ganoderma lucidum* caused to upregulate MVA related genes *FPS*, squalene synthase (*SQS*), or lanosterol synthase (*LS*), leading to develop the contents of ganoderic acid and intermediates, including squalene and lanosterol (Xu et al., 2012). Like *the effect of* overexpression of *BjHMGS1* in tomatoes, which significantly increased transcript levels of *FPS*, *SQS*, squalene epoxidase (*SQE*), and cycloartenol synthase (*CAS*; Liao et al., 2018), this study also exhibited that the overexpression of *PtHMGR* caused to upregulate the MVA related genes *AACT*, *MVK*, *FPS*, *MVD, and except HMGS. These results proved that* the MVA-related genes contribute to the biosynthesis of sesquiterpenes.

Overexpression of MEP-related gene *TwDXR* in *Tripterygium wilfordii* promoted the expressions of MVA-related genes *TwHMGS*, *TwHMGR*, *TwFPS*, and even MEP-related gene *TwGPPS* but demoted the expression of *TwDXS* (Zhang et al., 2018). The other study also reported that the overexpression of MEP-related gene *NtDXR1* in tobacco upregulated the transcript levels of eight MEP-related genes, indicating that the overexpression of *NtDXR1* led to an improved expression of MEP-related genes (Zhang et al., 2015). In *A. thaliana*, the overexpression of *DXR* revealed no influence on regulating *DXS* gene expression or enzyme accumulation, although overexpression of *DXR* promotes MEP-derived isoprenoids such as carotenoids, chlorophylls, and taxadiene (Carretero-Paulet et al., 2006). Overexpression of potato *DXS* in *A. thaliana* led to upregulate *GGPPS* and phytoene synthase (*PSY*; Henriquez et al., 2016). In this study, the overexpression of *PtDXR* upregulated the MEP-related genes. Controversy, the *PtDXR-OEs* revealed

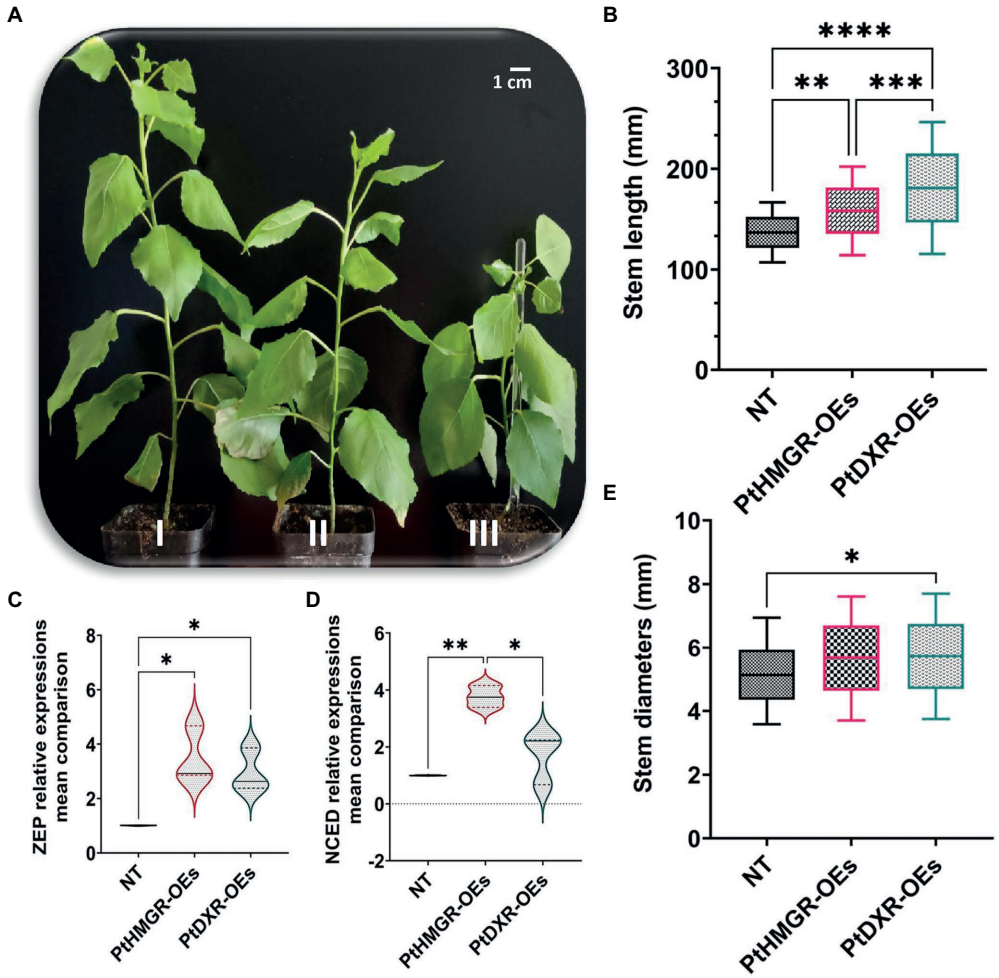


FIGURE 6
Phenotypic changes analyses resulted from the overexpression of *PtHMGR* and *PtDXR* effects on MVA and MEP pathways. (**A**, I), The *PtDXR* transgenic revealed a higher stem length than *PtHMGR*-OEs and NT poplars. (**A**, II), The *PtHMGR* transgenic presents more stem development than NT poplar. (**A**, III), NT poplar was used as a control; Scale bar represents 1cm. **(B)** The Box and Whisker mean comparison plot of stem lengths revealed significantly higher lengths of *PtDXR*-OEs than NT poplars compared with *PtHMGR*-OEs. *PtHMGR* transgenics also revealed significantly higher lengths than NT poplars. **(C,D)** The Violin mean comparison plots of *ZEP* and *NCED* relative expressions between *PtHMGR-and PtDXR*-OEs compared with NT poplars. **(E)** The Box and Whisker mean comparison plot of stem diameters revealed slightly more diameters between *PtDXR*-OEs and NT poplars. Stars reveal significant differences, *p <0.05, **p <0.01, ***p <0.001, ****p <0.0001.

significant downregulation of MVA-related genes. This study indicated that *PtHMGR* and *PtDXR* have important influences on MEP and MVA-related gene expression.

Overexpression of *PtHMGR* and *PtDXR* promotes the formation of GA_3, and carotenoids in plastids and the accumulation of TZR, IPA, and DCS in the cytoplasm

HMGR is a rate-limiting enzyme in the MVA pathway of plants and plays a critical role in controlling the flow of carbon within this metabolic pathway. The upregulation of *HMGR* significantly increases isoprenoid levels in plants. The overexpression of *HMGR* has been reported in several plants caused to promote the isoprenoid levels significantly. The heterologous expression of *Hevea brasiliensis HMGR1* in tobacco increased the sterol content and accumulated intermediate metabolites (Schaller et al., 1995). In addition, *the o*verexpression of *A. thaliana HMGR* in *Lavandula latifolia* increased the levels of sterols in the MVA and MEP-derived monoterpenes and sesquiterpenes (Munoz-Bertomeu et al., 2007). In addition, the overexpression *of Salvia miltiorrhiza SmHMGR* in hairy roots increased the MEP-derived diterpene tanshinone (Kai et al., 2011). In this study, ABA synthesis-related genes (*NCED1*, *NCED3*, *NCED6*, *ZEP1*, *ZEP2*, and *ZEP3*) and the contents of GA_3 and carotenoids were upregulated in *PtHMGR-OEs*. These findings suggest that the overexpression *of HMGR* may indirectly affect the biosynthesis of MEP-derived isoprenoids, including GA_3 and carotenoids. The accumulation of MVA-derived isoprenoids, including TZR,

IPA, and DCS, was significantly elevated in *PtHMGR-OEs*, indicating that overexpression *of PtHMGR* directly influences the biosynthesis of MVA-related isoprenoids.

DXR is the rate-limiting enzyme in the MEP pathway and is an essential regulatory step in the cytoplasmic metabolism of isoprenoid compounds (Aharoni et al., 2005). It has been shown that the overexpression of *DXR* in *Mentha piperita* promotes the synthesis of monoterpenes in the oil glands and increases the production of essential oil yield by 50% (Mahmoud and Croteau, 2001). Another study exhibited that overexpression of *DXR* from *Synechocystis* sp. strain PCC6803 in tobacco resulted in increasing the β-carotene, chlorophyll, antheraxanthin, and lutein (Hasunuma et al., 2008). In addition, it has been shown that *dxr* mutants of *A. thaliana* revealed a lack of GAs, ABA, and photosynthetic pigments (REF57; Xing et al., 2010) and showed pale sepals and yellow inflorescences (Xing et al., 2010). This study exhibits that the overexpression of *PtDXR* positively affects the accumulation of GA_3 and carotenoids, besides upregulation of MEP-related genes *DXS, DXR, IDI, HDS*, *HDR*, *MCT*, *CMK*, *GPS*, and *GPPS*.

Cross-talk exists between MVA and MEP pathways in excess of IPP and DMAPP

Although the substrates of MVA and MEP pathways differ, there are common intermediates, like IPP and DMAPP (Figure 6). Blocking only the MVA or the MEP pathway, respectively, does not entirely prevent the biosynthesis of terpenes in the cytoplasm or plastids, indicating that MVA and MEP pathway products can be transported and/or move between cell compartments (Aharoni et al., 2003, 2004; Gutensohn et al., 2013). For example, it has been shown that the transferring of IPP from the chloroplast to the cytoplasm was observed through ^{13}C labeling (Ma et al., 2017). In addition, segregation between the MVA and MEP pathways is limited and might exchange some metabolites over the plastid membrane (Laule et al., 2003). Some studies have applied the clustered, regularly interspaced short palindromic repeats (CRISPR) technology to reconstruct the lycopene synthesis pathway and control the flow of carbon in the MEP and MVA pathways (Kim et al., 2016). Results showed that the expressions of MVA-related genes were reduced by 81.6%, but the lycopene yield was significantly increased. By analyzing gene expression levels and metabolic outcomes in *PtHMGR-OEs* and *PtDXR-OEs*, we discovered a correlation between MVA and MEP-related genes among the derived products, which are not restricted to cross-talk between IPP and DMAPP (Figure 6).

The overexpression of *PtDXR* affected the transcript levels of MEP-related genes and the contents of MEP-derived isoprenoids, including GA_3 and carotenoids. The weakened expressions of MVA-related genes reduce the yields of MVA-derived isoprenoids (including CS) but increase the TZR, IPA, and DCS contents. This study hypothesizes that IPP and DMAPP produced by the MEP pathway pass the cytoplasm to compensate for the lack of IPP and DMAPP of the MVA pathway to synthesize MVA-derived products.

The analysis of the overexpression of *PtHMGR* in *PtHMGR-OEs* exhibited higher transcript levels of MVA-related genes *AACT*, *MVK*, *FPS, MVD*, and except *HMGS* and MEP-related genes *DXS*, *DXR*, *HDS*, *HDS*, *HDR*, *IDI*, *GPS,* and *GPPS* than NT poplars, and proved that the overexpression of MVA related gene *PtHMGR* involves in regulating of both MEP and MVA pathways. These results demonstrate that cytosolic *HMGR* overexpression expanded plastidial *GPP-and GGPP*-derived products, such as GA_3 and carotenoids, through cross-talk between MVA and MEP pathways. These results illustrate that regulation in the expression of MEP and MVA-related genes affects the MVA and MEP-derived isoprenoids.

The advanced insights in regulating MVA and MEP pathways in poplars improve the knowledge about these pathways in Arabidopsis, tomato, and rice. Altogether, these results discover that manipulating *PtDXR and PtHMGR* is a novel strategy to control poplar isoprenoids.

The overexpression of *PtHMGR* and *PtDXR* affects cross-talk between MVA and MEP pathways, resulting in plant growth and developments

ABA and GA_3 have been proved that perform essential functions in cell division, shoot growth, and flower induction (Xing et al., 2016). It has also been shown that the cytokinin TZR, a variety of phytohormones, performs important functions in the development and growth processes in shoots (Sakakibara, 2006; Abul et al., 2010). This study proved that the existing cross-talk between MVA and MEP pathways would be impacted positively by the overexpression of *PtHMGR* and *PtDXR*, leading to enhanced plant growth and development.

Conclusion

Isoprenoid compounds, which have a variety of structures and properties, play a critical role in the survival and growth of plants. In this study, *PtHMGR* showed positive effects on the accumulation of MVA-derived isoprenoids and MEP-derived substances, such as ABA, GA_3, carotenoids, and GRs, which are essential hormones for controlling growth and stress responses. *PtDXR* also exhibited an association with regulating MVA and MEP-related genes and derived isoprenoids.

Data availability statement

The data presented in the study are deposited in NCBI (https://www.ncbi.nlm.nih.gov/nuccore/XM_006384809.2/) repository, accession number XM_006384809.2. The original contributions presented in the study are included in the article/Supplementary material.

Author contributions

AM and HW conceived, planned, and coordinated the project, performed data analysis, wrote the draft, and finalized the manuscript. AM supervised this project. BP validated and contributed to data analysis and curation, revised and finalized the manuscript. MG-Z, FR, AK-P, and TJ reviewed and edited the manuscript. QZ, LY, and XZ coordinated, contributed to data curation, finalized, and funded this research. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by the National Natural Science Foundation of China (nos. 32001326 and 31971682), the Zhejiang Provincial Natural Science Foundation of China (nos. LZ19C160001 and LQ21C160003), the National Key Program on Transgenic Research (2018ZX08020002), and the Talent Research Foundation of Zhejiang A&F University (no. 2019FR055).

Conflict of interest

The authors declare that this research was conducted without any commercial or financial relationships construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.968780/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Amino acid sequences alignment of PtHMGR protein and other known HMGR proteins. The high similarities between *Populus* species and the other species are exhibited. The name of species and also their accession numbers are presented. The HMG-CoA and NADPH binding domains are indicated in red rectangular.

SUPPLEMENTARY FIGURE 2
Construction of a phylogenetic tree based on the HMGR sequences of various species. *HMGR* has been isolated from 35 species via blast on NCBI and aligned, followed by a phylogenetic tree. Results reveal a highly conserved *HMGR* through all the species. The gray rectangular shows similar *HMGR1* via *Populus* species located on one root, followed by a red dashed rectangle to show the organisms and accession numbers. The names of organisms and corresponding accession numbers are presented.

SUPPLEMENTARY FIGURE 3
Confirming of transformation of *PtHMGR-OEs*. **(A)** Schematic of constructed cassette transformed into the *PtHMGR* transgenic poplars. **(B)** PCR identification of *PtHMGR* in *PtHMGR-OEs* and NT poplars. Lane M: 15K molecular mass marker (TransGen, China); lane 1: genome DNA from NT as negative control; lanes 2–9: genome DNA from *PtHMGR-OE* lines. **(C)** qRT-PCR identifies the transcript levels of *PtHMGR* in *PtHMGR-OEs* and NT poplars. Three independent experiments were performed; Stars reveal significant differences, $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

SUPPLEMENTARY FIGURE 4
MEP and MVA-related genes analyses in *PtHMGR-* and *PtDXR-OEs* poplars. **(A)** MVA-related genes *AACT*, *HMGS*, *HMGR*, *MVK*, *MVD,* and *FPS* are affected by the overexpression of *PtHMGR*. **(B)** MVA-related genes *AACT*, *HMGS*, *HMGR*, *MVK*, *MVD,* and *FPS* are affected bythe overexpression of *PtDXR*. **(C)** MEP-related genes *DXS*, *MCT*, *CMK*, *HDS*, *HDR*, *IDI, GPS, GPPS, and DXR* are affected by the overexpression of *PtHMGR*. **(D)** MEP-related genes *DXS*, *DXR*, *MCT*, *CMK*, *HDS*, *HDR*, *IDI*, *GPS,* and *GPPSare* affected by the overexpression of *PtHMGR*. *PtActin* was used as an internal reference in all repeats; $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$; Three independent replications were performed in this experiment.

SUPPLEMENTARY FIGURE 5
Chromatogram analyses of GA_3 standards via HPLC-MS/MS. The chromatogram of standard GA_3 at **(A)** 0.1, **(B)** 0.2, **(C)** 0.5, **(D)** 2, **(E)** 5, **(F)** 20, **(G)** 50, and **(H)** 200 ng/ml concentrations. **(I)** Equations for the GA_3 standard curves.

SUPPLEMENTARY FIGURE 6
Chromatogram analyses of TZR standards via HPLC-MS/MS. The chromatogram of standard TZR at **(A)** 0.1, **(B)** 0.2, **(C)** 0.5, **(D)** 2, **(E)** 5, **(F)** 20, **(G)** 50, and **(H)** 200 ng/ml concentrations. **(I)** Equations for the TZR standard curves.

SUPPLEMENTARY FIGURE 7
Chromatogram analyses of IPA standards via HPLC-MS/MS. The chromatogram of standard IPA at **(A)** 0.2, **(B)** 0.5, **(C)** 2, **(D)** 5, **(E)** 20, **(F)** 50, and **(G)** 200 ng/ml concentrations. **(H)** Equations for the IPA standard curves.

SUPPLEMENTARY FIGURE 8
Chromatogram analyses of DCS standards via HPLC-MS/MS. The chromatogram of standard DCS at **(A)** 0.5, **(B)** 2, **(C)** 10, **(D)** 20, and **(E)** 50 ng/ml concentrations. **(F)** Equations for the DCS standard curves.

SUPPLEMENTARY FIGURE 9
Chromatogram analyses of CS standards via HPLC-MS/MS. The chromatogram of standard CS at **(A)** 0.5, **(B)** 5, **(C)** 10, **(D)** 20, and **(E)** 50 ng/ml concentrations. **(F)** Equations for the CS standard curves.

References

Abul, Y., Menendez, V., Gomez-Campo, C., Revilla, M. A., Lafont, F., and Fernandez, H. (2010). Occurrence of plant growth regulators in Psilotum nudum. *J. Plant Physiol.* 167, 1211–1213. doi: 10.1016/j.jplph.2010.03.015

Aharoni, A., Giri, A. P., Deuerlein, S., Griepink, F., Kogel, W. J., Verstappen, F. W., et al. (2003). Terpenoid metabolism in wild-type and transgenic Arabidopsis plants. *Plant Cell* 15, 2866–2884. doi: 10.1105/tpc.016253

Aharoni, A., Giri, A. P., Verstappen, F. W., Bertea, C. M., Sevenier, R., Sun, Z., et al. (2004). Gain and loss of fruit flavor compounds produced by wild and cultivated strawberry species. *Plant Cell* 16, 3110–3131. doi: 10.1105/tpc.104.023895

Aharoni, A., Jongsma, M. A., and Bouwmeester, H. J. (2005). Volatile science? Metabolic engineering of terpenoids in plants. *Trends Plant Sci.* 10, 594–602. doi: 10.1016/j.tplants.2005.10.005

Bohlmann, J., and Keeling, C. I. (2008). Terpenoid biomaterials. *Plant J.* 54, 656–669. doi: 10.1111/j.1365-313X.2008.03449.x

Boutte, Y., and Grebe, M. (2009). Cellular processes relying on sterol function in plants. *Curr. Opin. Plant Biol.* 12, 705–713. doi: 10.1016/j.pbi.2009.09.013

Bouvier, F., Rahier, A., and Camara, B. (2005). Biogenesis, molecular regulation and function of plant isoprenoids. *Prog. Lipid Res.* 44, 357–429. doi: 10.1016/j.plipres.2005.09.003

Carretero-Paulet, L., Cairo, A., Botella-Pavia, P., Besumbes, O., Campos, N., Boronat, A., et al. (2006). Enhanced flux through the methylerythritol 4-phosphate pathway in Arabidopsis plants overexpressing deoxyxylulose 5-phosphate reductoisomerase. *Plant Mol. Biol.* 62, 683–695. doi: 10.1007/s11103-006-9051-9

Cordoba, E., Salmi, M., and Leon, P. (2009). Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J. Exp. Bot.* 60, 2933–2943. doi: 10.1093/jxb/erp190

Cowan, A. K., Moore-Gordon, C. S., Bertling, I., and Wolstenholme, B. N. (1997). Metabolic control of avocado fruit growth (Isoprenoid growth regulators and the reaction catalyzed by 3-Hydroxy-3-Methylglutaryl coenzyme A Reductase). *Plant Physiol.* 114, 511–518. doi: 10.1104/pp.114.2.511

Dai, Z., Cui, G., Zhou, S. F., Zhang, X., and Huang, L. (2011). Cloning and characterization of a novel 3-hydroxy-3-methylglutaryl coenzyme A reductase gene from salvia miltiorrhiza involved in diterpenoid tanshinone accumulation. *J. Plant Physiol.* 168, 148–157. doi: 10.1016/j.jplph.2010.06.008

Devappa, R. K., and Rakshit, S. K., and and Dekker, R. F. (2015). Forest biorefinery: potential of poplar phytochemicals as value-added co-products. *Biotechnol. Adv.*, 33, 681–716. doi: 10.1016/j.biotechadv.2015.02.012

Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V., et al. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* 27, 753–759. doi: 10.1038/nbt.1557

Enfissi, E. M., Fraser, P. D., Lois, L. M., Boronat, A., Schuch, W., and Bramley, P. M. (2005). Metabolic engineering of the mevalonate and non-mevalonate isopentenyl diphosphate-forming pathways for the production of health-promoting isoprenoids in tomato. *Plant Biotechnol. J.* 3, 17–27. doi: 10.1111/j.1467-7652.2004.00091.x

Gershenzon, J., and Dudareva, N. (2007). The function of terpene natural products in the natural world. *Nat. Chem. Biol.* 3, 408–414. doi: 10.1038/nchembio.2007.5

Ghirardo, A., Wright, L. P., Bi, Z., Rosenkranz, M., Pulido, P., Rodriguez-Concepcion, M., et al. (2014). Metabolic flux analysis of plastidic isoprenoid biosynthesis in poplar leaves emitting and nonemitting isoprene. *Plant Physiol.* 165, 37–51. doi: 10.1104/pp.114.236018

Gutensohn, M., Orlova, I., Nguyen, T. T., Davidovich-Rikanati, R., Ferruzzi, M. G., Sitrit, Y., et al. (2013). Cytosolic monoterpene biosynthesis is supported by plastid-generated geranyl diphosphate substrate in transgenic tomato fruits. *Plant J.* 75, 351–363. doi: 10.1111/tpj.12212

Hain, R., Reif, H. J., Krause, E., Langebartels, R., Kindl, H., Vornam, B., et al. (1993). Disease resistance results from foreign phytoalexin expression in a novel plant. *Nature* 361, 153–156. doi: 10.1038/361153a0

Hasunuma, T., Takeno, S., Hayashi, S., Sendai, M., Bamba, T., Yoshimura, S., et al. (2008). Overexpression of 1-Deoxy-D-xylulose-5-phosphate reductoisomerase gene in chloroplast contributes to increment of isoprenoid production. *J. Biosci. Bioeng.* 105, 518–526. doi: 10.1263/jbb.105.518

Hemmerlin, A., and Harwood, J. L., and and Bach, T. J. (2012). A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Prog. Lipid Res.*, 51, 95–148. doi: 10.1016/j.plipres.2011.12.001

Henriquez, M. A., Soliman, A., Li, G., Hannoufa, A., Ayele, B. T., and Daayf, F. (2016). Molecular cloning, functional characterization and expression of potato (Solanum tuberosum) 1-deoxy-d-xylulose 5-phosphate synthase 1 (StDXS1) in response to Phytophthora infestans. *Plant Sci.* 243, 71–83. doi: 10.1016/j.plantsci.2015.12.001

Henry, L. K., Gutensohn, M., Thomas, S. T., Noel, J. P., and Dudareva, N. (2015). Orthologs of the archaeal isopentenyl phosphate kinase regulate terpenoid production in plants. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10050–10055. doi: 10.1073/pnas.1504798112

Huchelmann, A., Gastaldo, C., Veinante, M., Zeng, Y., Heintz, D., Tritsch, D., et al. (2014). S-carvone suppresses cellulase-induced capsidiol production in Nicotiana tabacum by interfering with protein isoprenylation. *Plant Physiol.* 164, 935–950. doi: 10.1104/pp.113.232546

Kai, G., Xu, H., Zhou, C., Liao, P., Xiao, J., Luo, X., et al. (2011). Metabolic engineering tanshinone biosynthetic pathway in salvia miltiorrhiza hairy root cultures. *Metab. Eng.* 13, 319–327. doi: 10.1016/j.ymben.2011.02.003

Kang, S. M., Min, J. Y., Kim, Y. D., Karigar, C. S., Kim, S. W., Goo, G. H., et al. (2009). Effect of biotic elicitors on the accumulation of bilobalide and ginkgolides in Ginkgo biloba cell cultures. *J. Biotechnol.* 139, 84–88. doi: 10.1016/j.jbiotec.2008.09.007

Kim, M. S., Haney, M. J., Zhao, Y., Mahajan, V., Deygen, I., Klyachko, N. L., et al. (2016). Development of exosome-encapsulated paclitaxel to overcome MDR in cancer cells. *Nanomedicine* 12, 655–664. doi: 10.1016/j.nano.2015.10.012

Kim, Y. J., Lee, O. R., Ji, Y. O., and Jang, M. G., and and Yang, D. C. (2014). Functional analysis of HMGR encoding genes in triterpene saponin-producing Panax ginseng Meyer. *Plant Physiol.*, 165, 373–387. doi: 10.1104/pp.113.222596

Kim, M. J., Noh, M. H., Woo, S., Lim, H. G., and Jung, G. Y. (2019). Enhanced lycopene production in Escherichia coli by expression of two MEP pathway enzymes from vibrio sp. *Dhg. Catalysts* 9:1003. doi: 10.3390/catal9121003

Kirby, J., and Keasling, J. D. (2009). Biosynthesis of plant isoprenoids: perspectives for microbial engineering. *Annu. Rev. Plant Biol.* 60, 335–355. doi: 10.1146/annurev.arplant.043008.091955

Kong, L. Y., and Tan, R. X. (2015). Artemisinin, a miracle of traditional Chinese medicine. *Nat. Prod. Rep.* 32, 1617–1621. doi: 10.1039/c5np00133a

Laule, O., Fürholz, A., Chang, H. S., Zhu, T., Wang, X., Heifetz, P. B., et al. (2003). Cross-talk between cytosolic and plastidial pathways of isoprenoid biosynthesis in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 6866–6871. doi: 10.1073/pnas.1031755100

Liao, Z. H., Chen, M., Gong, Y. F., Miao, Z. Q., Sun, X. F., and Tang, K. X. (2006). Isoprenoid biosynthesis in plants: pathways, genes, regulation and metabolic engineering. *J. Biol. Sci.* 6, 209–219. doi: 10.3923/jbs.2006.209.219

Liao, P., Chen, X., Wang, M., Bach, T. J., and Chye, M. L. (2018). Improved fruit alpha-tocopherol, carotenoid, squalene and phytosterol contents through manipulation of Brassica juncea 3-Hydroxy-3-Methylglutaryl-Coa Synthase1 in transgenic tomato. *Plant Biotechnol. J.* 16, 784–796. doi: 10.1111/pbi.12828

Liao, P., Hemmerlin, A., Bach, T. J., and Chye, M. L. (2016). The potential of the mevalonate pathway for enhanced isoprenoid production. *Biotechnol. Adv.* 34, 697–713. doi: 10.1016/j.biotechadv.2016.03.005

Lu, X. M., Hu, X. J., Zhao, Y. Z., Song, W. B., Zhang, M., Chen, Z. L., et al. (2012). Map-based cloning of zb7 encoding an IPP and DMAPP synthase in the MEP pathway of maize. *Mol. Plant* 5, 1100–1112. doi: 10.1093/mp/sss038

Lu, X., Tang, K., and Li, P. (2016). Plant metabolic engineering strategies for the production of pharmaceutical Terpenoids. *Front. Plant Sci.* 7:1647. doi: 10.3389/fpls.2016.01647

Ma, D., Li, G., Zhu, Y., and Xie, D. Y. (2017). Overexpression and suppression of Artemisia annua 4-Hydroxy-3-Methylbut-2-enyl Diphosphate Reductase 1 gene (AaHDR1) differentially regulate Artemisinin and Terpenoid biosynthesis. *Front. Plant Sci.* 8:77. doi: 10.3389/fpls.2017.00077

Ma, Y., Yuan, L., Wu, B., Li, X., Chen, S., and Lu, S. (2012). Genome-wide identification and characterization of novel genes involved in terpenoid biosynthesis in salvia miltiorrhiza. *J. Exp. Bot.* 63, 2809–2823. doi: 10.1093/jxb/err466

Mahmoud, S. S., and Croteau, R. B. (2001). Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase. *Proc. Natl. Acad. Sci. U. S. A.* 98, 8915–8920. doi: 10.1073/pnas.141237298

Mahmoud, S. S., and Croteau, R. B. (2002). Strategies for transgenic manipulation of monoterpene biosynthesis in plants. *Trends Plant Sci.* 7, 366–373. doi: 10.1016/s1360-1385(02)02303-8

Malkin, R., and Niyogi, K. (2000). *Photosynthesis*. Rockville: American Society of Plant Physiologists.

Maurey, K., Wolf, F., and Golbeck, J. (1986). 3-Hydroxy-3-Methylglutaryl coenzyme A Reductase activity in Ochromonas malhamensis: A system to study the relationship between enzyme activity and rate of steroid biosynthesis. *Plant Physiol.* 82, 523–527. doi: 10.1104/pp.82.2.523

Morrone, D., Lowry, L., Determan, M. K., Hershey, D. M., Xu, M., and Peters, R. J. (2010). Increasing diterpene yield with a modular metabolic engineering system in E. coli: comparison of MEV and MEP isoprenoid precursor pathway engineering. *Appl. Microbiol. Biotechnol.* 85, 1893–1906. doi: 10.1007/s00253-009-2219-x

Movahedi, A., Zhang, J., Amirian, R., and Zhuge, Q. (2014). An efficient agrobacterium-mediated transformation system for poplar. *Int. J. Mol. Sci.* 15, 10780–10793. doi: 10.3390/ijms150610780

Movahedi, A., Zhang, J., Sun, W., Mohammadi, K., Almasi Zadeh Yaghuti, A., Wei, H., et al. (2018). Functional analyses of PtRDM1 gene overexpression in poplars and evaluation of its effect on DNA methylation and response to salt stress. *Plant Physiol. Biochem.* 127, 64–73. doi: 10.1016/j.plaphy.2018.03.011

Movahedi, A., Zhang, J. X., Yin, T. M., and Qiang, Z. G. (2015). Functional analysis of two orthologous NAC genes, CarNAC3, and CarNAC6 from Cicer arietinum, involved in abiotic stresses in poplar. *Plant Mol. Biol. Rep.* 33, 1539–1551. doi: 10.1007/s11105-015-0855-0

Munoz-Bertomeu, J., Sales, E., Ros, R., Arrillaga, I., and Segura, J. (2007). Up-regulation of an N-terminal truncated 3-hydroxy-3-methylglutaryl CoA reductase enhances production of essential oils and sterols in transgenic Lavandula latifolia. *Plant Biotechnol. J.* 5, 746–758. doi: 10.1111/j.1467-7652.2007.00286.x

Opitz, S., Nes, W. D., and Gershenzon, J. (2014). Both methylerythritol phosphate and mevalonate pathways contribute to biosynthesis of each of the major isoprenoid classes in young cotton seedlings. *Phytochemistry* 98, 110–119. doi: 10.1016/j.phytochem.2013.11.010

Perreca, E., Rohwer, J., Gonzalez-Cabanelas, D., Loreto, F., Schmidt, A., Gershenzon, J., et al. (2020). Effect of drought on the Methylerythritol 4-phosphate

(MEP) pathway in the isoprene emitting conifer Picea glauca. *Front. Plant Sci.* 11:546295. doi: 10.3389/fpls.2020.546295

Ren, D., Liu, Y., Yang, K. Y., Han, L., Mao, G., Glazebrook, J., et al. (2008). A fungal-responsive MAPK cascade regulates phytoalexin biosynthesis in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 105, 5638–5643. doi: 10.1073/pnas.0711301105

Roberts, S. C. (2007). Production and engineering of terpenoids in plant cell culture. *Nat. Chem. Biol.* 3, 387–395. doi: 10.1038/nchembio.2007.8

Rodriguez-Concepcion, M. (2010). Supply of precursors for carotenoid biosynthesis in plants. *Arch. Biochem. Biophys.* 504, 118–122. doi: 10.1016/j.abb.2010.06.016

Russell, D. W., and Davidson, H. (1982). Regulation of cytosolic HMG-CoA reductase activity in pea seedlings: contrasting responses to different hormones, and hormone-product interaction, suggest hormonal modulation of activity. *Biochem. Biophys. Res. Commun.* 104, 1537–1543. doi: 10.1016/0006-291x(82)91426-7

Sakakibara, H. (2006). Cytokinins: activity, biosynthesis, and translocation. *Annu. Rev. Plant Biol.* 57, 431–449. doi: 10.1146/annurev.arplant.57.032905.105231

Schaller, H., Grausem, B., Benveniste, P., Chye, M. L., Tan, C. T., Song, Y. H., et al. (1995). Expression of the Hevea brasiliensis (H.B.K.) Mull. Arg. 3-Hydroxy-3-Methylglutaryl-coenzyme A Reductase 1 in tobacco results in sterol overproduction. *Plant Physiol.* 109, 761–770. doi: 10.1104/pp.109.3.761

Song, X., Yu, X., Hori, C., Demura, T., Ohtani, M., and Zhuge, Q. (2016). Heterologous overexpression of poplar SnRK2 genes enhanced salt stress tolerance in Arabidopsis thaliana. *Front. Plant Sci.* 7:612. doi: 10.3389/fpls.2016.00612

Thulasiram, H. V., Erickson, H. K., and Poulter, C. D. (2007). Chimeras of two isoprenoid synthases catalyze all four coupling reactions in isoprenoid biosynthesis. *Science* 316, 73–76. doi: 10.1126/science.1137786

Tong, Y., Su, P., Zhao, Y., Zhang, M., Wang, X., Liu, Y., et al. (2015). Molecular cloning and characterization of DXS and DXR genes in the Terpenoid biosynthetic pathway of Tripterygium wilfordii. *Int. J. Mol. Sci.* 16, 25516–25535. doi: 10.3390/ijms161025516

Vaccaro, M., Malafronte, N., Alfieri, M., Tommasi, N., and Leone, A. (2014). Enhanced biosynthesis of bioactive abietane diterpenes by overexpressing AtDXS or AtDXR genes in Salvia sclarea hairy roots. *Plant Cell Tiss. Org.* 119, 65–77. doi: 10.1007/s11240-014-0514-4

van Schie, C. C., Haring, M. A., and Schuurink, R. C. (2006). Regulation of terpenoid and benzenoid production in flowers. *Curr. Opin. Plant Biol.* 9, 203–208. doi: 10.1016/j.pbi.2006.01.001

Wang, H., Nagegowda, D. A., Rawat, R., Bouvier-Nave, P., Guo, D., Bach, T. J., et al. (2012). Overexpression of Brassica juncea wild-type and mutant HMG-CoA synthase 1 in Arabidopsis up-regulates genes in sterol biosynthesis and enhances sterol production and stress tolerance. *Plant Biotechnol. J.* 10, 31–42. doi: 10.1111/j.1467-7652.2011.00631.x

Wei, Y., Mohsin, A., Hong, Q., Guo, M., and Fang, H. (2018). Enhanced production of biosynthesized lycopene via heterogenous MVA pathway based on chromosomal multiple position integration strategy plus plasmid systems in *Escherichia coli*. *Bioresour. Technol.* 250, 382–389. doi: 10.1016/j.biortech.2017.11.035

Wille, A., Zimmermann, P., Vranova, E., Furholz, A., Laule, O., Bleuler, S., et al. (2004). Sparse graphical Gaussian modeling of the isoprenoid gene network in Arabidopsis thaliana. *Genome Biol.* 5:R92. doi: 10.1186/gb-2004-5-11-r92

Wright, L. P., Rohwer, J. M., Ghirardo, A., Hammerbacher, A., Ortiz-Alcaide, M., Raguschke, B., et al. (2014). Deoxyxylulose 5-phosphate synthase controls flux through the Methylerythritol 4-phosphate pathway in Arabidopsis. *Plant Physiol.* 165, 1488–1504. doi: 10.1104/pp.114.245191

Xie, Z., Kapteyn, J., and Gang, D. R. (2008). A systems biology investigation of the MEP/terpenoid and shikimate/phenylpropanoid pathways points to multiple levels of metabolic control in sweet basil glandular trichomes. *Plant J.* 54, 349–361. doi: 10.1111/j.1365-313X.2008.03429.x

Xing, S., Miao, J., Li, S., Qin, G., Tang, S., Li, H., et al. (2010). Disruption of the 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR) gene results in albino, dwarf and defects in trichome initiation and stomata closure in Arabidopsis. *Cell Res.* 20, 688–700. doi: 10.1038/cr.2010.54

Xing, L., Zhang, D., Zhao, C., Li, Y., Ma, J., An, N., et al. (2016). Shoot bending promotes flower bud formation by miRNA-mediated regulation in apple (Malus domestica Borkh.). *Plant Biotechnol. J.* 14, 749–770. doi: 10.1111/pbi.12425

Xu, C., Wei, H., Movahedi, A., Sun, W., Ma, X., Li, D., et al. (2019). Evaluation, characterization, expression profiling, and functional analysis of DXS and DXR genes of Populus trichocarpa. *Plant Physiol. Biochem.* 142, 94–105. doi: 10.1016/j.plaphy.2019.05.034

Xu, J. W., Xu, Y. N., and Zhong, J. J. (2012). Enhancement of ganoderic acid accumulation by overexpression of an N-terminally truncated 3-hydroxy-3-methylglutaryl coenzyme A reductase gene in the basidiomycete Ganoderma lucidum. *Appl. Environ. Microbiol.* 78, 7968–7976. doi: 10.1128/AEM.01263-12

Yamaguchi, S., and Kamiya, Y., and and Nambara, E. (2018). Regulation of ABA and GA levels during seed development and germination in Arabidopsis. *Annu. Plant Rev.*, 27, 224–247. doi: 10.1002/9781119312994.apr0283

Zhang, K. K., Fan, W., Huang, Z. W., Chen, D. F., Yao, Z. W., Li, Y. F., et al. (2019). Transcriptome analysis identifies novel responses and potential regulatory genes involved in 12-deoxyphorbol-13-phenylacetate biosynthesis of Euphorbia resinifera. *Ind. Crop Prod.* 135, 138–145. doi: 10.1016/j.indcrop.2019.04.030

Zhang, J., Li, J., Liu, B., Zhang, L., Chen, J., and Lu, M. (2013). Genome-wide analysis of the Populus Hsp90 gene family reveals differential expression patterns, localization, and heat stress responses. *BMC Genomics* 14:532. doi: 10.1186/1471-2164-14-532

Zhang, J., Movahedi, A., Sang, M., Wei, Z., Xu, J., Wang, X., et al. (2017). Functional analyses of NDPK2 in Populus trichocarpa and overexpression of PtNDPK2 enhances growth and tolerance to abiotic stresses in transgenic poplar. *Plant Physiol. Biochem.* 117, 61–74. doi: 10.1016/j.plaphy.2017.05.019

Zhang, H., Niu, D., Wang, J., Zhang, S., Yang, Y., Jia, H., et al. (2015). Engineering a platform for photosynthetic pigment, hormone and Cembrane-related Diterpenoid production in Nicotiana tabacum. *Plant Cell Physiol.* 56, 2125–2138. doi: 10.1093/pcp/pcv131

Zhang, Y., Zhao, Y., Wang, J., Hu, T., Tong, Y., Zhou, J., et al. (2018). Overexpression and RNA interference of TwDXR regulate the accumulation of terpenoid active ingredients in Tripterygium wilfordii. *Biotechnol. Lett.* 40, 419–425. doi: 10.1007/s10529-017-2484-1

frontiers | Frontiers in Plant Science

TYPE Systematic Review
PUBLISHED 14 December 2022
DOI 10.3389/fpls.2022.1064176

OPEN ACCESS

EDITED BY
Maria F. Drincovich,
Centro de Estudios Fotosintéticos y Bioquímicos (CEFOBI), Argentina

REVIEWED BY
Agnieszka Zienkiewicz,
Nicolaus Copernicus University in Toruń, Poland
Hubert Schaller,
UPR2357 Institut de biologie moléculaire des plantes (IBMP), France

*CORRESPONDENCE
Xiao-Hong Yu
xhyu@bnl.gov
John Shanklin
shanklin@bnl.gov

SPECIALTY SECTION
This article was submitted to Plant Systems and Synthetic Biology, a section of the journal Frontiers in Plant Science

RECEIVED 07 October 2022
ACCEPTED 28 November 2022
PUBLISHED 14 December 2022

CITATION
Cai Y, Yu X-H and Shanklin J (2022) A toolkit for plant lipid engineering: Surveying the efficacies of lipogenic factors for accumulating specialty lipids.
Front. Plant Sci. 13:1064176.
doi: 10.3389/fpls.2022.1064176

A toolkit for plant lipid engineering: Surveying the efficacies of lipogenic factors for accumulating specialty lipids

Yingqi Cai[1], Xiao-Hong Yu[1,2*] and John Shanklin[1*]

[1]Biology Department, Brookhaven National Laboratory, Upton, NY, United States, [2]Department of Biochemistry and Cell Biology, Stony Brook University, Stony Brook, NY, United States

Plants produce energy-dense lipids from carbohydrates using energy acquired *via* photosynthesis, making plant oils an economically and sustainably attractive feedstock for conversion to biofuels and value-added bioproducts. A growing number of strategies have been developed and optimized in model plants, oilseed crops and high-biomass crops to enhance the accumulation of storage lipids (mostly triacylglycerols, TAGs) for bioenergy applications and to produce specialty lipids with increased uses and value for chemical feedstock and nutritional applications. Most successful metabolic engineering strategies involve heterologous expression of lipogenic factors that outperform those from other sources or exhibit specialized functionality. In this review, we summarize recent progress in engineering the accumulation of triacylglycerols containing - specialized fatty acids in various plant species and tissues. We also provide an inventory of specific lipogenic factors (including accession numbers) derived from a wide variety of organisms, along with their reported efficacy in supporting the accumulation of desired lipids. A review of previously obtained results serves as a foundation to guide future efforts to optimize combinations of factors to achieve further enhancements to the production and accumulation of desired lipids in a variety of plant tissues and species.

KEYWORDS

lipid engineering, fatty acid, triacylglycerol, lipid droplet, specialty fatty acid, specialty lipid, lipogenic factor

Introduction

All organisms can convert carbohydrates into fatty acids (FAs), the building blocks of both phospholipids for membrane synthesis and triacylglycerols (TAGs) for carbon and energy storage. Some organisms including plants have evolved specialized lipogenic factors to accumulate large quantities of TAGs or produce specialty FAs. Bio-based

TAGs, also known as storage lipids, contain more than twice the energy of carbohydrates, making them a sustainable energy-dense source of biofuels (Ohlrogge and Chapman, 2011; Singh et al., 2021). Specialty lipids containing high levels of specialty FAs can serve as feedstocks for jet fuel, nutraceuticals, and industrial products because of their distinct physical and functional properties (Dyer et al., 2008; Park et al., 2021). However, natural sources of these lipids are limited and therefore are not sufficient to meet growing demand. Plants use carbon and energy acquired from photosynthesis to synthesize FAs and accumulate TAGs and thus represent a renewable and economically viable platform for lipid production. General conservation of lipid synthesis across kingdoms makes it possible to engineer agronomic plants for the production and accumulation of desired lipids by inter-species heterologous expression of many lipogenic factors.

In plant cells, FAs are synthesized from acetyl-coenzyme A (CoA) in plastids (Figure 1; Ohlrogge and Browse, 1995; Li-Beisson et al., 2013). The heteromeric acetyl-CoA carboxylase (ACCase) catalyzes the conversion of acetyl-CoA to malonyl-CoA, the first committed step in FA synthesis. With acetyl-CoA serving as the starting unit, the acyl chain is extended by the FA synthase complex (FAS) through sequential condensation of two-carbon units from malonyl-acyl carrier protein (ACP). FAs reaching a certain chain length (typically C16 or C18) are released from ACP by fatty acyl thioesterases (FAT)A/B and exported from plastids. These FAs then enter the acyl-CoA pool in the endoplasmic reticulum (ER), where they are further modified and incorporated into glycerolipids (Figure 1). ER-localized FA elongase (FAE) can add additional two-carbon units to acyl-CoA to further elongate FAs. The acyl chains esterified to phosphatidylcholine (PC) undergo modifications to introduce double bond(s) and functional groups to FAs and the modified FAs re-enter the acyl-CoA pool through acyl-editing, a dynamic acyl exchange between PC and the acyl-CoA pool. TAGs can be assembled *via* sequential acylation of glycerol-3-phosphate (G3P) with acyl-CoA as the acyl donor catalyzed by glycerol-3-phosphate acyltransferase (GPAT), lysophosphatidyl acyltransferase (LPAT), phosphatidic acid phosphatase (PAP), and diacylglycerol:acyl-CoA acyltransferase (DGAT) (reviewed in Li-Beisson et al., 2013; Xu and Shanklin, 2016). Alternatively, PC can serve as the acyl donor for acylation of diacylglycerol (DAG) to form TAG by phospholipid:diacylglycerol acyltransferase (PDAT) (Dahlqvist et al., 2000). In addition, FAs esterified to PC may enter the TAG pool through the conversion of PC to DAG and subsequently to TAG by phosphatidylcholine: diacylglycerol cholinephosphotransferase (PDCT) or phospholipase C (PLC) (Wang, 2001; Lu et al., 2009). TAGs

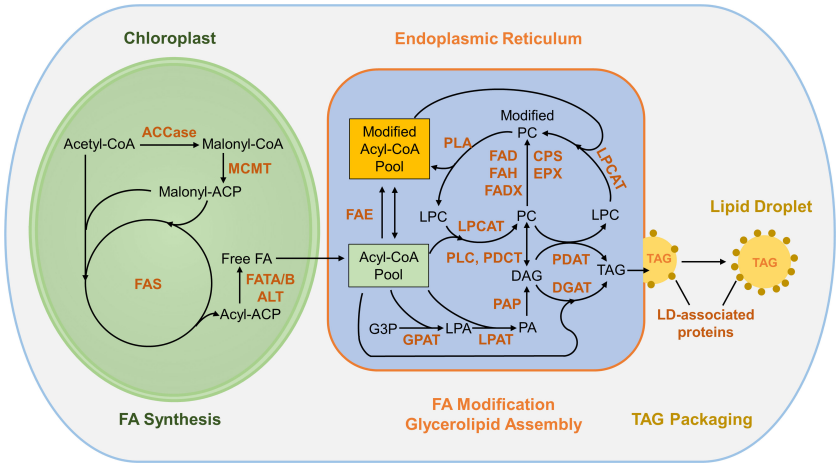


FIGURE 1
Overview of FA synthesis and TAG assembly in plants. The conversion of acetyl-CoA to malonyl-CoA by ACCase is the first committed step in fatty acid synthesis in plastids. With acetyl-CoA serving as the starting unit, the acyl chain is extended by sequential condensation of two-carbon units from malonyl-ACP by FAS complex. FAs exported from plastids enter the acyl-CoA pool in the ER and can be incorporated into PC, where acyl groups are modified and re-enter the acyl-CoA pool thorough acyl editing. The Kennedy pathway incorporates acyl-CoA into glycerolipids *via* sequential acylation of G3P by GPAT, LPAT and DGAT. TAG can be synthesized through acyl-CoA-dependent (DGAT converting DAG and acyl-CoA to TAG) and acyl-CoA-independent (PDAT synthesizing TAG from DAG and PC) pathways. TAGs are packaged into LDs and protected by LD-associated proteins. ACCase, acetyl-CoA carboxylase; ACP, acyl carrier protein; CoA, coenzyme A; MCMT, malonyl-CoA: ACP malonyltransferase; FAS, fatty acid synthase; FATA/B, fatty acyl thioesterase A/B; ALT, acyl-lipid thioesterase; FAE, fatty acid elongase; GPAT, glycerol-3-phosphate acyltransferase; LPAT, lysophosphatidyl acyltransferase; PAP, phosphatidic acid phosphatase; DGAT, diacylglycerol:acyl-CoA acyltransferase; PDAT: phospholipid:diacylglycerol acyltransferase; PLC, phospholipase C; PDCT, phosphatidylcholine:diacylglycerol cholinephosphotransferase; LPCAT, lysophosphatidylcholine acyltransferase; PLA, phospholipase A; CPS, cyclopropane synthase; EPX, epoxygenase; FAD, fatty acid desaturase; FAH; fatty acyl hydroxylase; FADX, fatty acid conjugase; G3P, glycerol-3-phosphate; LPA, lysophosphatidic acid; PA, phosphatidic acid; DAG, diacylglycerol; PC, phosphatidylcholine; LPC, lysophosphatidylcholine; TAG, triacylglycerol.

synthesized within the bilayer of the ER membrane are packaged into cytosolic lipid droplets (LDs) by coordination of LD-related proteins including SEIPIN, LDAP, LDIP, Oleosin, Caleosin, and Steroleosin (reviewed by Chapman et al., 2012 and Pyc et al., 2017).

In plants, most TAGs are synthesized and stored in seeds, serving as a major reserve of carbon and energy for seed germination and seedling establishment. In contrast, plant vegetative tissues usually contain trace amounts of TAGs despite their high capacity for FA synthesis. Studies of lipid engineering in plants increasingly focus on 1) producing value-added specialty lipids in seeds of oilseed crops (e.g., *Camelina sativa* and *Brassica napus*) by introducing specialized lipogenic factors to increase the value of seed oil (Xu et al., 2020; Yuan and Li, 2020) and 2) enhancing the accumulation of TAGs in vegetative tissues of high-biomass crops (e.g., *Nicotiana tabacum*, *Sorghum bicolor*, and *Saccharum* spp. *Hybrids*) by overexpressing TAG-enhancing lipogenic factors to increase the overall lipid yield on a per plant and unit land area basis (Vanhercke et al., 2019b; Park et al., 2021). Almost all these lipid engineering approaches require heterologous expression of lipogenic factors from other organisms. In this review, we present a survey of work evaluating the efficacy of lipogenic factors from various organisms in engineering lipids in selected target plant species. Specifically, we provide a list of lipogenic factors that exhibit specialized functional features in FA synthesis and modification (Section I and Table 1), glycerolipid assembly (Section II and Table 2) and LD biogenesis (Section III and Table 3) and highlight recent progress in optimizing combinations of lipogenic factors for enhanced production of desired lipids, and discuss challenges and future opportunities for lipid engineering in plants. It can be difficult to correlate the effects of expressed genes on lipid metabolism, especially those discussed in review articles, with their precise coding sequences. In this work we have included accession numbers wherever possible to address and remedy this issue.

Section I. Producing specialized FAs by tailoring FA synthesis and modification

Lipogenic factors involved in FA synthesis and modification determine the diversity of FAs with respect to their carbon chain lengths, degree of unsaturation, and addition of a variety of functional groups, which determine their physical properties and potential industrial uses. Typical FAs found conserved in the plant kingdom range from 16 to 18 carbons in length and contain 0 to 3 double bounds at Δ^9, Δ^{12} and Δ^{15} positions (i.e., counting relative to the carboxyl group). In contrast to these "common" FAs, some FAs with shorter or longer chain lengths, additional double bonds, double bond(s) at different position(s), or functional groups at specific locations along the carbon chain are found in specific groups of plant species or non-plant organisms, and thus are referred to as specialty FAs. The structural properties of these specialty FAs make them promising feedstocks for biofuels, industrial products and nutraceuticals. To increase the value of plant lipids, engineering strategies involving the heterologous expression of lipogenic factors related to FA synthesis and modification have been developed to produce specialty FAs in both seed and vegetative tissues of domesticated plant species (Park et al., 2021).

In this section, we describe efforts to evaluate enzymes in the FA biosynthesis and modification pathway that are responsible for producing the following well-studied types of specialty FAs. Medium-chain FAs result from the action of FA thioesterases that release acyl chains from acyl carrier protein (ACP). Hydroxy, epoxy and conjugated FAs arise from the action of enzymes that evolved from the Δ^{12}-oleic FA desaturase 2 (FAD2) class of integral membrane desaturases (Shanklin and Cahoon, 1998) which act primarily on oleic acid esterified to PC. Omega-7 monounsaturated FAs with a double bound at the ω^7 position (i.e., counting relative to the methyl end of FAs) can be produced by Δ^9-acyl-ACP or Δ^9-acyl-CoA desaturase with high specificity for 16:0-ACP or 16:0-CoA, respectively (Bondaruk et al., 2007; Nguyen et al., 2010). Very-long-chain PUFAs arise from the action of multiple desaturases and elongases, and their engineering represents a tour-de-force in heterologous expression and pathway optimization (Napier et al., 2019). The last example is the addition of a cyclopropyl group across the double bond in oleic acid by cyclopropane synthase, a class of enzymes present in plants and prokaryotes (Bao et al., 2002; Bao et al., 2003). In section II and III, we summarize the approaches to incorporate these specialized FAs into TAGs and subsequently package them into LDs.

Medium-chain fatty acids

Medium-chain FAs (MCFAs) include FAs of 8-14 carbons in lengths, generated by the hydrolysis of FA from acyl carrier protein between the C8 and C14 stages of elongation *via* variants of FATB with defined chain length specificities (Figure 1). Lipids containing MCFAs are naturally produced in palm kernel (*Elaeis guineensis*), coconut (*Cocos nucifera*), and cuphea genus (*Cuphea pulcherrima*, *Cuphea viscosissima*, *Cuphea palustris*, *Cuphea hookeriana*), and these plants derived MCFAs serve as potential feedstocks for jet fuel and industrial products such as cosmetics and detergents (Dyer et al., 2008; Kallio et al., 2014; Park et al., 2021). To engineer the production of MCFAs in oilseed crops, FATB variants that specifically hydrolyze C8-C14 FAs from acyl-ACPs were isolated from California bay (*Umbellularia californica*) and Cuphea and expressed in *Arabidopsis thaliana*, *Camelina sativa*, and *Brassica napus*. Heterologous expression of *U. californica FATB1* produced

TABLE 1 List of specialized lipogenic factors involved in FA synthesis and modification used for lipid engineering in plants.

Lipogenic Factor	Origin Species	Accession No.	Target Species	Effects of Heterologous Expression on Lipid Metabolism	Reference
FATB1	*U. californica*	M94159	*B. napus*	Produced seed oil containing up to 50% MCFA (lauric acid 12:0).	Eccleston et al., 1996; Voelker et al., 1996
			A. thaliana	Produced MCFA (lauric acid 12:0) accounted for up to 37% and 43% of seed oil in wild-type and *aae15/16* mutant backgrounds, respectively.	Tjellström et al., 2013
	U. californica	M94159 Q41635	*C. sativa*	Produced MCFA (lauric acid, 12:0 and myristic acid, 14:0) accounted for 21% of seed oil.	Kim et al., 2015b
	C. viscosissima	AEM72522	*C. sativa*	Produced MCFA (C8-C14) accounted for 15% of seed oil.	Kim et al., 2015b
	C. pulcherrima	AGG79283	*C. sativa*	Produced MCFA (myristic acid, 14:0) accounted for 1.6% of seed oil.	Kim et al., 2015b
FATB2	*C. palustris*	AAC49180	*A. thaliana*	Produced MCFA (myristic acid, 14:0) accounted for up to 39% and 42% of seed oil in wild-type and *aae15/16* mutant backgrounds, respectively.	Tjellström et al., 2013
			C. sativa	Produced MCFA (myristic acid, 14:0) accounted for 24% of seed oil.	Kim et al., 2015b
	C. hookeriana	AAC49269	*B. napus*	Produced MCFA (caprylic acid, 8:0; capric acid, 10:0; lauric acid, 12:0) accounted for up to 40% of seed oil.	Dehesh et al., 1996
			A. thaliana	Produced MCFA (caprylic acid, 8:0; capric acid, 10:0) accounted for up to 22% and 25% of seed oil in wild-type and *aae15/16* mutant backgrounds, respectively.	Tjellström et al., 2013
	C. hookeriana	AAC49269	*C. sativa*	Produced MCFA (C8-C14) accounted for 12.6% of seed oil.	Kim et al., 2015b
FATB3	*C. pulcherrima*	KC675178	*A. thaliana*	Produced MCFA (caprylic acid, 8:0; capric acid, 10:0) accounted for up to 6% and 12% of seed oil in wild-type and *aae15/16* mutant backgrounds, respectively.	Tjellström et al., 2013
			C. sativa	Produced MCFA (C8-C14) accounted for 2.9% of seed oil.	Kim et al., 2015b
			C. sativa	Produced MCFA (myristic acid, 14:0) accounted for 7.5% of seed oil.	Kim et al., 2015b
ALT1	*A. thaliana*	NM_103226 At1g35290	*C. sativa*	Produced MCFA (lauric acid, 12:0 and myristic acid, 14:0) accounted for up to 3.5% of seed oil.	Kalinger et al., 2021
	A. thaliana	NM_103226 At1g35290	*N. benthamiana*	Produced approximately 50 nmol MCFA (lauric acid, 12:0 and myristic acid, 14:0) per gram leaf fresh weight.	Kalinger et al., 2021
ALT4	*A. thaliana*	NM_001334359 At1g68280	*C. sativa*	Produced approximately 1% MCFA (caproic acid, 6:0; caprylic acid, 8:0; capric acid, 10:0; and myristic acid, 14:0) in seed oil.	Kalinger et al., 2021
			N. benthamiana	Produced approximately 53 nmol MCFA (caproic acid, 6:0) per gram leaf fresh weight.	Kalinger et al., 2021
FATB2 KAS4	*C. hookeriana*	AAC49269 AF060519	*B. napus*	Increased MCFA by 30-40% in seed oil as compared to that of plants expressing *ChFATB2* alone.	Dehesh et al., 1998
FATB1 KAS4	*C. palustris* *C. hookeriana*	U38188 AF060519	*B. napus*	Increased MCFA by 40% in seed oil as compared to that of plants expressing *ChFATB2* alone.	Dehesh et al., 1998
FAH	*C. purpurea*	EU661785	*A. thaliana*	Produced hydroxy FA (ricinoleic and densipolic) up to 25% of seed oil in the Arabidopsis fad2/fae1 mutant.	Meesapyodsuk and Qiu, 2008
FAH12	*R. communis*	U22378	*A. thaliana*	Produced hydroxy FA (ricinoleic, densipolic, lesquerolic, and auricolic acids) accounted for up to 19% of seed oil in wild-type, *fad2/fae1*, *fad3*, and *fad3/fae1* plants.	Broun and Somerville, 1997; Smith et al., 2003
			C. sativa	Produced hydroxy FA to approximately 15% of seed oil in the wild-type background.	Aryal and Lu, 2018
FAH12-1	*H. benghalensis*	KC533767	*A. thaliana*	Produced up to 21% hydroxy FA in seed oil of the *fad2/fae1* mutant.	Zhou et al., 2013
FAH12-2	*H. benghalensis*	KC533768	*A. thaliana*	Produced up to 18% hydroxy FA in seed oil of the *fad2/fae1* mutant.	Zhou et al., 2013

(Continued)

TABLE 1 Continued

Lipogenic Factor	Origin Species	Accession No.	Target Species	Effects of Heterologous Expression on Lipid Metabolism	Reference
EPX	*C. palaestina*	Y16283	*A. thaliana*	Produced epoxy FA accounted for up to 6.2% of seed oil, which was further increased to 21% when co-expressed with *CpFAD2* in the *fad3/fae1* mutant.	Singh et al., 2001; Zhou et al., 2006
			G. hirsutum	Produced epoxy FA accounted for 17% of seed oil when co-expressed with *CpFAD2*.	Zhou et al., 2006
	E. lagascae	AF406732	*N. tabacum*	Produced epoxy FA accounted for 15% of total FA in calli.	Cahoon et al., 2002
			G. max	Produced epoxy FA accounted for 8% of total FA in somatic embryos.	Cahoon et al., 2002
	S. laevis	AY462108	*A. thaliana*	Produced 2.4% epoxy FA (vernolic acid) in seed oil.	Hatanaka et al., 2004
			P. hybrida	Produced epoxy FA (vernolic acid) accounted for 0.5% of total lipids when transiently expressed in leaves.	Li et al., 2010
			G. max	Produced 8% epoxy FA (vernolic acid) in seed oil.	Li et al., 2010
	V. galamensis	N/A	*N. benthamiana*	Produced epoxy FA accounted for 8.7% of total leaf lipids, which was further increased to 13.1% when co-expressed with *VgFAD2*.	Sun et al., 2022
FADX	*M. charantia*	AF18252	*G. max*	Produced conjugated FA (eleostearic and parinaric acids) accounted for up to 18% of total FA in somatic embryos.	Cahoon et al., 1999
			A. thaliana	Produced eleostearic acid accounted for approximately 13% of total seed FA in the *fad3/fae1* mutant.	Cahoon et al., 2006
	I. balsamina	AF182520	*G. max*	Produced conjugated FA (eleostearic and parinaric acids) accounted for up to 5% of total FA in somatic embryos.	Cahoon et al., 1999
	C. officinalis	AF310156	*G. max*	Produced calendic acid accounted for approximately 22% of total FA in somatic embryos.	Cahoon et al., 2006
			A. thaliana	Produced calendic acid accounted for approximately 15% of total seed FA in the *fad3/fae1* mutant.	Cahoon et al., 2006
	V. fordii	AF525535	*A. thaliana*	Produced eleostearic acid accounted for approximately 6% of total seed FA in the *fad3/fae1* mutant.	Cahoon et al., 2006
			A. thaliana	Produced approximately 2% eleostearic acid in leaf neutral lipids.	Yurchenko et al., 2017
Δ^9-AAD	*A. syriaca*	U60277	*A. thaliana*	Failed to produce detectable ω-7 FA in seed oil.	Bondaruk et al., 2007
Δ^9-AAD	*D. unguis-cati*	AF051134	*A. thaliana*	Produced approximately 28% and 9% ω-7 FAs in Arabidopsis and Brassica seed oil, respectively.	Bondaruk et al., 2007
Com25 (mutated Δ^9-AAD)	*R. communis*	N/A	*A. thaliana*	Resulted in accumulation of ω-7 FAs to 14% and 56% of seed oil when expressed in wild-type and fab1/fae1 backgrounds, respectively.	Nguyen et al., 2010
			C. sativa	Increased the content of ω-7 FAs to approximately 17% of seed oil.	Nguyen et al., 2015
SnD9D AnD9D	*S. nodorum* *A. nidulans*		*A. thaliana*	Produced ω-7 FAs accounted for approximately 24% of seed oil and further increased the level of ω-7 FAs to up to 71% of seed oil when co-expressed with Com 25 in fab1/fae1 seeds.	Nguyen et al., 2010
Com25 FAT5	*R. communis* *C. elegans*		*C. sativa*	Produced ω-7 FAs accounted for approximately 23% and 65% of seed oil in wild-type and fab1/fae1/fatb backgrounds, respectively.	Nguyen et al., 2015
CPS	*S. foetida*	AF470622	*A. thaliana*	Produced a trace amount of CPA (~0.05% of total FA) in seeds of the *fad2/fae1* mutant.	Yu et al., 2011
	E. coli	M98330	*N. benthamiana*	Produced up to 3.7% CPA in total FA in leaves, which was increased to 11.8% when *NbFAD2* was silenced, and a novel C18:2CPA.	Okada et al., 2020
	E. coli	944811	*A. thaliana*	Produced substantial amounts of CPA (up to 9.1% of total FA) in seeds of the *fad2/fae1* mutant.	Yu et al., 2014
			C. sativa	Produced up to approximately 10% CPA in total seed FA.	Yu et al., 2018
CPS1	*G. hirsutum*	AY574036	*A. thaliana*	Produced detectable amounts of CPA (up to 1% of total FA) in seeds of the *fad2/fae1* mutant.	Yu et al., 2011

(Continued)

TABLE 1 Continued

Lipogenic Factor	Origin Species	Accession No.	Target Species	Effects of Heterologous Expression on Lipid Metabolism	Reference
			N. benthamiana	Produced up to 1% CPA in total FA in leaves, which was increased to 4.8% when *NbFAD2* was silenced.	Okada et al., 2020
Δ^{12}-DES Δ^{15}/ω^{3}-DES Δ^{6}-DES Δ^{6}-ELO Δ^{5}-DES Δ^{5}-ELO Δ^{4}-DES	*P. sojae* *P. infestans* *O. tauri* *P. patens* *T.* sp. *O. tauri* *O.* RCC809	EGZ11023 XP_002902599 XP_003082578 AAL84174 AAM09687 CAI58913 JGI: 40461	*C. sativa*	Represent an optimal combination of genes for EPA and DPA biosynthesis in oilseeds. Routinely produced EPA and DHA in excess of 20% total seed oil.	Han et al., 2020; Han et al., 2022

N/A, not available.

TABLE 2 List of specialized lipogenic factors involved in glycerolipid assembly used for lipid engineering in plants.

Lipogenic Factor	Origin Species	Accession No.	Target Species	Effects of Heterologous Expression on Lipid Metabolism	Reference
LPAT	*C. nucifera*	U29657	*B. napus*	Enabled efficient MCFA (lauric acid, 12:0) deposition at the *sn*-2 position of TAG, resulting in accumulation of lauric acid to over 50% of seed oil when co-expressed with *UcFATB1*.	Knutzon et al., 1999
	C. nucifera	Q42670 U29657	*C. sativa*	Increased lauric acid (12:0) and myristic acid (14:0) in seed oil and at the *sn*-2 position of TAG when co-expressed with *UcFATB1* and *CpFATB2*, respectively.	Kim et al., 2015b
	V. galamensis	N/A	*N. benthamiana*	Increased the level of epoxy FA from 8.7% to 16.7% when co-expressed with *VgEPX*.	Sun et al., 2022
	S. foetida	KC894726	*A. thaliana*	Enriched CPA in glycerolipids and increased CPA accumulation up to 35% of total seed FA when co-expressed with *EcCPS* in the *fad2/fae1* mutant.	Yu et al., 2014
			C. sativa	Enriched CPA in glycerolipids and increased CPA levels up to 18% of total seed FA when co-expressed with *EcCPS* in the *fad2/fae1* mutant.	Yu et al., 2018
LPAT2	*R. communis*	EU591533	*A. thaliana*	Slightly increased the level of hydroxy FA in seed oil in the *fae1* mutant expressing *RcFAH12*.	Shockey et al., 2019
	V. fordii	MH823254	*A. thaliana*	Significantly increased eleostearic acid content in seed oil in the *fad3/fae1* mutant expressing *VfFADX*.	Shockey et al., 2019
	C. viscosissima	ALM22867	*C. sativa*	Enabled deposition of capric acid (10:0) at the *sn*-2 position of TAG.	Kim et al., 2015a
LPAT2a	*C. pulcherrima*	ALM22869	*C. sativa*	Enabled deposition of capric acid (10:0) at the *sn*-2 position of TAG.	Kim et al., 2015a
LPATB	*C. pulcherrima*	ALM22873	*C. sativa*	Enabled deposition of myristic acid (14:0) but not capric acid (10:0) at the *sn*-2 position of TAG.	Kim et al., 2015a
LPCAT	*V. galamensis*	N/A	*N. benthamiana*	Increased the level of epoxy FA from 8.7% to 19.4% when co-expressed with *VgEPX*.	Sun et al., 2022
PDCT	*R. communis*	EQ973818	*A. thaliana*	Enriched hydroxy FA in DAG and TAG, increased hydroxy FA levels to nearly 20% of seed oil when co-expressed with *RcFAH12* in the wild-type background, and partially restored the decreased seed oil content caused by *RcFAH12* expression.	Hu et al., 2012
	L. chinensis	KU926346	*C. sativa*	Enhanced the transfer of CPA from PC to DAG and led to a 57% increase in CPA accumulation in TAG when co-expressed with *EcCPS* relative to expressing *EcCPS* alone in the *fad2/fae1* mutant.	Yu et al., 2019
PLCL1	*R. communis*	XM_002523576	*C. sativa*	Enriched hydroxy FA in TAG, increased hydroxy FA levels to 22% of seed oil when co-expressed with *RcFAH12*.	Aryal and Lu, 2018
PDAT1A (PDAT1-2)	*R. communis*	NM_001323733	*A. thaliana*	Channeled hydroxy FA into TAG and increased hydroxy FA to 27% of seed oil when co-expressed with *RcFAH12* in the *fae1* mutant.	van Erp et al., 2011
DGAT1	*C. pulcherrima*	KU055625	*C. sativa*	Enriched MCFA (capric acid, 10:0) in TAG and increased capric acid content to 14.5% of seed oil when co-expressed with *CvFATB1*.	Iskandarov et al., 2017
	V. galamensis	EF653277	*P. hybrida*	Resulted in a 2-fold increase in epoxy FA in leaves co-expressing *VgDGAT1* and *SlEPX* relative to expressing *SlEPX* alone.	Li et al., 2010
			G. max	Increased the accumulation of epoxy FA to 15% of seed oil when co-expressed with *SlEPX*.	Li et al., 2010

(Continued)

TABLE 2 Continued

Lipogenic Factor	Origin Species	Accession No.	Target Species	Effects of Heterologous Expression on Lipid Metabolism	Reference
	C. ellipsoidea	KT779429	*A. thaliana*	Increased seed oil content by 8–37%.	Guo et al., 2017
			B. napus	Increased seed oil content by 12–18%.	Guo et al., 2017
DGAT2	*R. communis*	EU391592	*A. thaliana*	Enhanced the incorporation of hydroxy FA into TAG and increased the level of hydroxy FA to approximately 30% of seed oil when co-expressed with *RcFAH12* in the *fae1* mutant.	Burgal et al., 2008; Shockey et al., 2019
	V. galamensis	FJ652577	*P. hybrida*	Resulted in a 6-fold increase in epoxy FA in leaves co-expressing *VgDGAT2* and *SlEPX* relative to expressing *SlEPX* alone.	Li et al., 2010
			G. max	Increased the accumulation of epoxy FA to 26% of seed oil when co-expressed with *SlEPX*.	Li et al., 2010
	V. fordii	DQ356682	*A. thaliana*	Redirected eleostearic acids from phospholipids to TAGs, increased eleostearic acid to approximately 12% of neutral lipids in leaves and mitigated the negative growth effects caused by *FADX* expression. No significant increase in eleostearic acid was observed in seeds.	Yurchenko et al., 2017; Shockey et al., 2019
	M. musculus	BC043447	*N. benthamiana*	Increased TAG contents by 20-fold when transiently expressed in leaves.	Cai et al., 2019
DGAT2-2	*C. esculentus*	N/A	*N. tabacum*	Increased TAG contents in leaves to 5.5% DW, which is 7.2-fold and 1.7-fold higher than that in wild-type leaves and leaves expressing *AtDGAT1*, respectively. Increased the proportion of oleic acid in leaf lipids.	Gao et al., 2021
DGAT5 (DGTT5)	*N. oceanica*	KY273672	*N. benthamiana*	Increased TAG contents by 2-fold when transiently expressed in leaves.	Zienkiewicz et al., 2017
			A. thaliana	Increased TAG contents by 6-fold in leaves and increased seed oil content by 50%.	Zienkiewicz et al., 2017
DGAT (DAcT)	*E. alatus*	GU594061	*A. thaliana*	Resulted in accumulation of acTAG up to 40% of total TAG in seed oil.	Durrett et al., 2010
			C. sativa	Produced an average of 52% acTAG in seed oil.	Alkotami et al., 2021
	E. fortunei	MF06125	*C. sativa*	Produced an average of 72% acTAG in seed oil.	Alkotami et al., 2021
LPAT, DGAT (DAcT)	*C. nucifera* *E. alatus*	Q42670 GU594061	*C. sativa*	Produced about 15% acTAG with MCFA in seeds expressing *UcFATB1* and with silenced endogenous *DGAT1* and *PDAT1*.	Bansal et al., 2018
LPAT2, DGAT1	*C. viscosissima* *C. pulcherrima*	ALM22867 KU055625	*C. sativa*	Enriched MCFA (capric acid, 10:0) in TAG and increased capric acid content to 23.7% of seed oil, which is higher than that in plants expressing these enzymes individually.	Iskandarov et al., 2017
LPAT2, DGAT2	*V. fordii*	MH823254 DQ356682	*A. thaliana*	Increased the content of eleostearic acids to nearly 30% of seed oil in the *fad3/fae1* mutant expressing *VfFADX*.	Shockey et al., 2019
LPAT2, DGAT2	*R. communis*	EU591533 EU391592	*A. thaliana*	Produced a higher level of hydroxy FA (up to 30% of seed oil) in the *fae1* mutant expressing *RcFAH12*, compared to expressing *RcLPAT2* or *RcDGAT2* alone.	Shockey et al., 2019
GPAT9, LPAT2, DGAT2	*R. communis*	EU391594 EU591533 EU391592	*A. thaliana*	Adding *RcGPAT9* to the combination of *RcLPAT2* and *RcDGAT2* did not further increased hydroxy FA content.	Shockey et al., 2019
GPAT9, LPAT2, PDAT1A	*R. communis*	NP_001310690 NP_001310679 NM_001323733	*A. thaliana*	Produced tri-hydroxy-TAG, increased hydroxy FA to 34% of seed oil, and restored seed oil content to wild-type level when co-expressed with *RcFAH12* in the *fae1* mutant.	Lunn et al., 2019
DGAT2, LPCAT, PDAT1-2, PDCT	*R. communis*	EU391592 KC540908 NM_001323733 EQ973818	*A. thaliana*	Produced hydroxy FA to approximately 25% and 31% of seed oil in the wild-type and *fae1* backgrounds, respectively, when co-expressed with *RcFAH12*.	Park et al., 2022

N/A, not available.

TABLE 3 List of specialized lipogenic factors involved in LD biogenesis used for lipid engineering in plants.

Lipogenic Factor	Origin Species	Accession No.	Target Species	Effects of Heterologous Expression on Lipid Metabolism	Reference
OLE	*S. indicum*	AAD42942	*S. tuberosum*	Increased TAG contents in leaves and tubers (3.3% TAG of DW) when combined with other lipogenic factors.	Liu et al., 2017
			N. tabacum	Increased TAG contents in leaves, stems, and roots when combined with other lipogenic factors.	Vanhercke et al., 2014 Vanhercke et al., 2017
			S. bicolor	Increase TAG (8.4% of DW) and total lipid (9.9% of DW) contents in leaves when combined with other lipogenic factors.	Vanhercke et al., 2019a
	R. communis	N/A	*A. thaliana*	Increased hydroxy FA from 18% to 22% of seed oil in Arabidopsis expressing *RcFAH12*.	Lu et al., 2006
Cys-OLE	*S. indicum*	N/A	*A. thaliana*	Enhanced the accumulation of lipids in leaves to a higher level compared with the wild-type SiOLE.	Winichayakul et al., 2013
FIT2	*M. musculus*	BAE37420	*A. thaliana*	Increased the number and size of LDs in leaves and enhanced lipid accumulation in both leaves and seeds.	Cai et al., 2017
			N. benthamiana	Promoted LD proliferation and increased levels of neutral lipids in leaves.	Cai et al., 2017
FSP27	*M. musculus*	NM_178373	*A. thaliana*	Increased the number and size of LDs in leaves, and elevated lipid contents in seeds.	Price et al., 2020
			N. benthamiana	Mediated LD fusion and increased the number and size of LDs in leaves.	Price et al., 2020

N/A, not available.

MCFAs primarily consisting of lauric acid (C12:0) up to 21%, 37%, and 50% of seed oil in Camelina, Arabidopsis, and *B. napus*, respectively (Eccleston et al., 1996; Voelker et al., 1996; Tjellström et al., 2013; Kim et al., 2015b). FATB variants from different Cuphea species showed different efficiencies and substrate chain length specificities when expressed in oilseed plants (Table 1). Expression of *FATB2* from *C. hookeriana* could produce MCFAs ranging from C8 to C14 with capric acid (C10:0) as the most abundant species accounted for approximate 12%, 22%, and 40% of total seed lipids in Camelina, Arabidopsis, and *B. napus*, respectively (Dehesh et al., 1996; Tjellström et al., 2013; Kim et al., 2015b). FATB1 from *C. viscosissima* produced 15% MCFAs with chain lengths varying from C8 to C14 in seed oil when expressed in Camelina, whereas FATB2 from *C. palustris* produced only myristic acid (C14:0) to 24% and 39% of seed oil in Camelina and Arabidopsis, respectively (Tjellström et al., 2013; Kim et al., 2015b). Relatively low levels of MCFAs (1.2%-7.5% of seed oil) were detected in seeds of Arabidopsis and Camelina expressing *FATB1*, *FATB3* or *FATB4* from *C. pulcherrima*, compared with FATBs from other Cuphea species. Based on data collected from Camelina, UcFATB1 and CpFATB2 seem to be the most effective FATB variants in producing MCFAs with UcFATB1 preferentially generating lauric acid (C12:0) and CpFATB2 exclusively producing myristic acid (C14:0). Furthermore, co-expression of *ChFATB2* or *CpFATB1* with a Cuphea medium-chain-specific 3-ketoacyl-ACP synthase (KAS4) that catalyzes the condensation of acyl-ACP with malonyl-ACP increased MCFA content by up to 40% in *B. napus* seed oil as compared with that of plants expressing *FATB* alone (Dehesh et al., 1998). Disruption of acyl-ACP synthetase (AAE15/16) that re-activates FAs released from acyl-ACP in Arabidopsis overexpressing Cuphea *FATB* further enhanced MCFA accumulation in seeds (Tjellström et al., 2013). In addition to the FAT-type thioesterases containing two "hotdog" folds (Mayer and Shanklin, 2005), another acyl-ACP thioesterase family, acyl-lipid thioesterase (ALT) with a single "hotdog" fold, is generally present in all classes of plants (Kalinger et al., 2020). Overexpression of Arabidopsis *ALT1* or *ALT4* in Camelina seeds and *Nicotiana benthamiana* leaves yielded C6–C14 MCFA at a relatively lower level (as much as 3.5% of seed oil) compared to the effective FATB isoforms (Table 1; Kalinger et al., 2021).

Hydroxy fatty acids

The hydroxylation of FAs is mediated by the action of FA hydroxylase (FAH), the first functionally divergent FAD2 homolog to be identified (Figure 1). That they are mechanistically related is evidenced by reports that as few as four substitutions between desaturase and hydroxylase can interconvert their functionality (Broun et al., 1998; Broadwater et al., 2002). FAs with hydroxyl groups attached to the acyl chain are useful feedstocks for the formulation of plastics and lubricants (Dyer et al., 2008). The major natural source of hydroxy FAs for industrial uses is castor bean (*Ricinus communis*), which accumulates 90% hydroxy FAs (mostly ricinoleic acid, C18:1-OH) in its seed oil. The enzyme responsible for synthesizing hydroxy FAs in castor is FAH12 (van de Loo et al., 1995). Heterologous expression of RcFAH12

in Arabidopsis led to accumulation of hydroxy FAs consisting of primarily ricinoleic acid accounted for up to 19% of seed oil in wild type, *fad2/fae1*, *fae1*, *fad3*, or *fad3/fae1* backgrounds (Broun and Somerville, 1997; Smith et al., 2003). Similarly, wild-type Camelina expressing RcFAH12 produced approximately 15% hydroxy FAs in seed oil (Aryal and Lu, 2018). Expression of *Hiptage benghalensis* hydroxylases *HbFAH12-1* and *HbFAH12-2* in Arabidopsis *fad2/fae1* mutant yielded up to 21% and 18% hydroxy FA, respectively, in seed oil (Zhou et al., 2013). In contrast to the plant derived FAH12, a FAH homolog isolated from a fungal pathogen, *Claviceps purpurea*, produced 25% hydroxy FAs in seed oil when expressed in the Arabidopsis *fad2/fae1* mutant (Meesapyodsuk and Qiu, 2008). While RcFAH12 can effectively produce hydroxy FAs in target plants and most plant engineering strategies to date have used RcFAH12 to synthesize hydroxy FAs, searching for a more effective FAH from other species to further enhance the accumulation of hydroxy FAs in bioengineered crops might be productive. Lesquerella (*Physaria fendleri*), a Brassicaceae species closely related to Arabidopsis and Camelina, produces about 50% of lesquerolic acid (C20:1-OH), an elongated form of ricinoleic acid, in its seed oil (Horn et al., 2016). Thus, Lesquerella represents a promising alternative industrial oilseed for HFA production, and specialized HFA-related factors in Lesquerella represent a promising source for engineering HFA accumulation in other crops (Horn et al., 2016; Chen et al., 2021). A recent study demonstrated the production of *erythro*-9,10-dihydroxystearate, a vicinal diol by an acyl-ACP desaturase variant *via* dioxygenase chemistry (Whittle et al., 2020). The identification of additional genes that are more efficient at vicinal diol production may facilitate large scale production of these compounds that are difficult to synthesize chemically.

Epoxy FAs

An epoxy group with its oxygen bridging between adjacent carbons of fatty acyl chains conveys unique chemical reactivity useful for the production of plastics, polymers, coatings, and glues. Epoxy FAs are enriched in seed oils of certain plant species belonging to the Asteraceae and Euphorbiaceae families (Cahoon et al., 2002). Interestingly, the biosynthesis of epoxy FAs in different plant species is catalyzed by different classes of epoxygenase (EPX) enzymes. Those responsible for epoxy FA biosynthesis in Asteraceae species such as *Crepis palaestina*, *Stokesia laevis*, and *Vernonia galamensis* are divergent forms of the FAD2 desaturase, whereas epoxygenases in Euphorbiaceae species such as *Euphorbia lagascae* are cytochrome P450 enzymes (Bafor et al., 1993; Liu et al., 1998; Cahoon et al., 2002). Despite the distinction of these two classes of EPX, heterologous expression of these enzymes in plants resulted in accumulation of similar levels of epoxy FAs (mostly vernolic acid, C18:1- Δ^{12}-epoxy FA) in seed oils (Table 1). Expression of the FAD2-like EPX coding genes from *C. palaestina* and *S. laevis* led to accumulation of approximately 2.4%-8% epoxy FAs in seed oils of Arabidopsis and soybean, and the cytochrome P450-type EPX from *E. lagascae* produced about 8% epoxy FAs in soybean somatic embryos (Singh et al., 2001; Cahoon et al., 2002; Hatanaka et al., 2004; Li et al., 2010). Providing the exogenous CpEPX with more linoleic acid (C18:2) substrate by disrupting *FAD3* and *FAE1* in Arabidopsis increased the levels of epoxy FAs to 8.6% of seed oil (Zhou et al., 2006). Previous studies suggested that heterologous expression of either type of EPX can reduce the accumulation of linoleic acid in target plants probably caused by decreased activity of the endogenous FAD2 enzyme, and co-expression of EPX with a typical FAD2 dramatically enhanced the production of epoxy FAs to 21% of seed oil in Arabidopsis *fad3/fae1* mutant (Singh et al., 2001; Cahoon et al., 2002; Zhou et al., 2006). In addition, epoxy FAs can also be engineered in non-seed tissues. Expression of *ElEPX* in tobacco (*Nicotiana tabacum*) calli produced epoxy FAs accounted for 15% of total lipids (Cahoon et al., 2002). Transient expression of *SlEPX* in petunia (*Petunia hybrida*) leaves or a FAD2-like EPX from *V. galamensis* in *N. benthamiana* leaves resulted in accumulation of 0.5% or 8.7% epoxy FAs in total leaf lipids, respectively (Li et al., 2010; Sun et al., 2022). Moreover, co-expression of *VgEPX* with *VgFAD2* increased the level of epoxy FAs to 13.1% of total lipids in *N. benthamiana* leaves (Sun et al., 2022). Collectively, divergent classes of EPX from different plant species seem to be equally effective in producing epoxy FAs in seeds and providing more linoleic acid by overexpressing a "typical" *FAD2* or disrupting *FAD3* and *FAE1* is critical for further increasing epoxy FA levels.

Omega-7 unsaturated fatty acids

Omega-7 unsaturated FAs (ω-7 FAs) are potential feedstocks for the production of octene, a high-demand industrial product used for polyethylene production (Nguyen et al., 2010). Some plants (e.g., milkweed [*Asclepias syriaca*] and cat's claw vine [*Doxantha unguis-cati*]) can naturally produce ω-7 FAs (e.g., palmitoleic acid 16:1Δ^9 and cis-vaccenic acid 18:1Δ^{11}) by Δ^9-acyl-ACP desaturase (AAD) with high specificity for 16:0-ACP (Cahoon et al., 1997; Cahoon et al., 1998). Heterologous expression of the milkweed 16:0-ACP desaturase in Arabidopsis failed to produce detectable ω-7 FA, while the Doxantha 16:0-ACP desaturase produced approximately 28% and 9% ω-7 FAs in Arabidopsis and Brassica seed oil, respectively (Bondaruk et al., 2007). An AAD variant with high specificity for converting 16:0-ACP to 16:1 Δ^9-ACP was selected from a pool of randomized mutants of castor AAD (Cahoon and Shanklin, 2000), and expression of this engineered enzyme (Com25) in Arabidopsis seeds resulted in accumulation of ω-7 FAs to 14% of seed oil (Nguyen et al.,

2010). Increasing the level of 16:0-ACP by silencing the 16:0-ACP elongase, β-ketoacyl-ACP synthase II (KASII/FAB1), in *fae1* mutant overexpressing Com25 further increased the content of ω-7 FAs to 56% of seed oil. Co-expressing Com25 with two fungal Δ^9-16:0-CoA desaturases from *Stagonospora nodorum* (SnD9D) and *Aspergillus nidulans* (AnD9D) in fab1/fae1 mutant increased ω-7 FA content to 71% of seed oil by desaturating saturated FAs after transfer from the plastid to the ER (Nguyen et al., 2010). A similar strategy co-expressing Com25 and a Δ^9-16:0-CoA desaturase from *Caenorhabditis elegans* (FAT5) in Camelina seeds with 16:0-ACP substrate pools increased by silencing genes encoding KASII/FAB1, FAE1, and16:0-ACP thioesterase (FATB) increased ω-7 FAs to 60-65% of seed oil (Nguyen et al., 2015).

Conjugated fatty acids

The FAD2 desaturases that produce conjugated FAs by converting Δ^9 and Δ^{12} double bonds to Δ^{11} and Δ^{13} double bonds are designated as FA conjugases (FADX) (Figure 1; Cahoon et al., 1999; Dyer et al., 2002). The higher oxidation rates of conjugated FAs relative to typical polyunsaturated FAs make them useful as drying agents in paints and inks. Conjugated FAs can also serve as health supplements as they have been reported to have fat-reducing and anticancer effects in animals (Lee et al., 2002; Dyer et al., 2008; Yuan et al., 2014). Natural sources of conjugated FAs include tung tree (*Vernicia fordii*), *Momordica charantia*, *Impatiens balsamina*, and *Calendula officinalis*, and genes encoding FADX enzymes have been isolated from these plant species and evaluated for their efficacy in producing conjugated FAs in model plants and oilseed crops. Ectopic expression of FADX coding genes from *I. balsamina*, *M. charantia*, and *C. officinalis* in somatic soybean (*Glycine max*) embryos resulted in production of conjugated FAs to approximately 5%, 18%, and 22% of total FAs, respectively (Cahoon et al., 1999). For engineering approaches carried out in Arabidopsis seeds, mutants with FA desaturase 3 (FAD3) and FAE1 disrupted are used to provide more substrates (linoleic acid) for FADX. Arabidopsis *fad3/fae1* mutants expressing *FADX* genes from *V. fordii*, *M. charantia*, and *C. officinalis* accumulated approximately 6%, 13%, and 15% conjugated FAs in seed oil, respectively (Cahoon et al., 1999; Cahoon et al., 2006). Recent attempts to engineer conjugated FAs in plant vegetative tissues by expressing VfFADX in Arabidopsis successfully produced conjugated FAs (eleostearic acid, 18:3, Δ9c, Δ11t, Δ13t) to 2% of total neutral lipids in leaves (Yurchenko et al., 2017). Among all FADX enzymes tested so far, CoFADX seems to be the most effective enzyme for producing high levels of conjugated FAs in both Arabidopsis and soybean, but different FADX orthologs produce different types of conjugated FAs. Conjugated FAs produced by CoFADX comprise exclusively calendic acid (18:3, Δ8t, Δ10t, Δ12c), while eleostearic acid is the primary conjugated FAs detected in transgenic plants expressing *VfFADX* or *McFADX*. In future efforts to engineer conjugated FAs, it will be important to select a FADX that produces high levels and desired types of conjugated FAs.

Very-long-chain polyunsaturated FAs

Very-long-chain polyunsaturated FAs (VLCPUFA) such as eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA) are FAs with 20 or 22 carbons and 4 to 6 double bonds. EPA and DHA are valuable nutraceuticals because of their beneficial roles in fetal neuronal system development and cardiovascular diseases prevention (Tocher et al., 2019). Marine fish oils are a major source of EPA, DHA and other ω^3-VLCPUFAs. However, the marine-sourced EPA and DHA are insufficient to meet the increasing demand for these FAs in human diet (Tocher et al., 2019). To develop sustainable alternative sources of EPA and DHA, in the past two decades, enormous research effort has been directed at engineering ω^3-VLCPUFAs in plant oils, and significant progress has been achieved in producing EPA and DHA in oilseed crops (Napier et al., 2019). There are two pathways that can produce VLCPUFAs from linoleic acids: 1) the conventional pathway including Δ^6-desaturase (DES), Δ^6-elongase (ELO), Δ^5-DES, Δ^5-ELO, and Δ^4-DES; and 2) the alternative pathway using Δ^9-ELO, Δ^8-DES, Δ^5-DES, Δ^5-ELO, and Δ^4-DES. Both pathways have been introduced into plants and successfully produced EPA and DHA in both seed and vegetative tissues (Qi et al., 2004; Wu et al., 2005; Wood et al., 2009; Petrie et al., 2010). It was later found that, besides the minimum five enzymes required for DHA biosynthesis in plants, introduction of a highly active Δ^{12}-DES from *Lachancea kluyveri* and a Δ^{15}/ω^3-DES with broad substrate specificity from *Pichia pastoris* could effectively increase the ratio of ω^3/ω^6 FAs and therefore the level of DHA in seed oil (Petrie et al., 2012; Petrie et al., 2014). To optimize the engineering strategy to produce high levels of EPA and DHA in plants, DES and ELO enzymes from a wide variety of species including algae, fungi, oomycetes, mosses, animals, and flowering plants were introduced into plants to test their efficacy in producing EPA and DHA in various plant species (Abbadi et al., 2004; Qi et al., 2004; Wu et al., 2005; Wood et al., 2009; Petrie et al., 2010; Petrie et al., 2012; Petrie et al., 2014; Han et al., 2020; Han et al., 2022). The optimal enzyme combination that achieved the highest levels of EPA and DHA (over 20%) in seed oil reported to date consists of a Δ^{12}-DES from *Phytophthora sojae*, Δ^{15}/ω^3-DES from *Phytophthora infestans*, Δ^6-DES from *Ostreococcus tauri*, Δ^6-ELO from *Physcomitrella patens*, Δ^5-DES from *Thraustochytrium* sp., Δ^5-ELO from *O. tauri*, and Δ^4-DES from *Ostreococcus* RCC809 (Han et al., 2020; Han et al., 2022).

Cyclopropane fatty acids

Cyclopropane FAs (CPAs), such as dihydrosterculic acid (9, 10-methylene octadecanoic acid) and lactobacillic acid (11, 12 methylene octadecanoic acid), are specialized FAs that contain a cyclopropane group (three-carbon carbocyclic ring) within the carbon chain. They are found in bacteria and certain plant species such as *Litchi chinensis*. The highly strained and reactive carbocyclic ring of CPA readily opens to form methyl-branched fatty acids, which exhibit unique physical and chemical properties such as low melting temperatures, resistance to oxidation, and propensity for self-polymerization, making CPAs suitable for application in lubricants, paints, and coatings (Carlsson et al., 2011). Ectopic expression of genes encoding cyclopropane synthase (CPS), the enzyme catalyzes the conversion of monounsaturated FAs to CPAs (Figure 1), successfully produced CPAs in plants that normally lack these compounds. Intriguingly, the CPS gene from *Escherichia coli* is more effective than CPS homologs isolated from plant species for synthesizing CPAs in plants (Table 1). Engineering of CPAs in Arabidopsis and *Camelina* seeds were carried out in mutant lines with reduced FA desaturase 2 (FAD2) and FA elongase 1 (FAE1) that accumulate increased levels of oleoyl substrates for CPA production. Expression of *CPS* genes cloned from *Sterculia foetida* and cotton (*Gossypium hirsutum*) in Arabidopsis *fad2/fae1* mutant lines led to the accumulation of CPA to between 0.05% and 1% of total seed FAs, respectively (Yu et al., 2011). In contrast, the expression of *E. coli* CPS resulted in nearly 10% CPA accumulation in seed oil in Arabidopsis and Camelina (Yu et al., 2014; Yu et al., 2018). Recently, researchers tested the ability of CPS enzymes to synthesize CPAs in plant vegetative tissues by transiently expressing cotton or *E. coli CPS* genes in *N. benthamiana* leaves and found that GhCPS1 produced up to 1% CPA of total leaf FAs while *EcCPS* expression led to the accumulation of CPA up to 3.7% of total lipids (Okada et al., 2020). The levels of CPAs in leaf lipids can be further elevated to 4.8% and 11.8% by silencing the expression of endogenous *NbFAD2* in leaves expressing *GhCPS1* and *EcCPS*, respectively (Okada et al., 2020). Therefore, future strategies to engineer CPA accumulation in plant seed and vegetative tissues will exploit EcCPS rather than plant CPS variants.

Section II. Optimizing lipid accumulation by channeling selected FA toward TAG

TAGs, also known as storage lipids, are the most abundant form of vegetable oils. They primarily accumulate in seeds to provide energy for seed germination and establishment. In plant vegetative tissues such as leaves, TAGs are barely detectable and serve primarily as transient intermediates for FAs removed from membrane lipids prior to their degradation (Xu and Shanklin, 2016). Accumulation of free FAs in cells and specialty FAs in membrane lipids can result in negative effects on plant growth. One reason for this is that they elicit feedback inhibition of FA synthesis *via* biotin attachment domain-containing protein (BADC), a negative regulator of ACCase (Yang et al., 2015; Salie et al., 2016; Zale et al., 2016; Keereetaweep et al., 2018; Yu et al., 2021). Channeling FA flux toward the TAG pool can reduce the accumulation of free FAs and remove specialty FAs from membrane lipids, thereby mitigating their negative growth effects and further enhancing the accumulation of lipids with desired acyl composition. Therefore, Lipogenic factors involved in glycerolipid assembly capable of accommodating specialty FA substrates represent critical targets for enhancing specialty FA-containing TAG accumulation in plants. In the following subsections, we summarize previous efforts to assess the efficacies of TAG-assembly-related enzymes for optimizing the accumulation of desired lipids.

Incorporating specialty FA into TAG

Incorporation of specialty FA into TAG requires specialized enzymes that can recognize the specialized FA for catalyzing multiple steps of TAG assembly (Figure 1). In the glycerolipid biosynthesis pathway, The ER-localized LPAT catalyzes the transfer of FAs from the acyl-CoA pool to LPA to form PA, which serves as a key intermediate for channeling FAs into TAGs and membrane lipids. Lysophosphatidylcholine acyltransferase (LPCAT) incorporates FAs into PC. PDCT and PLC catalyze the conversion of PC to DAG and thereby allow FAs esterified to PC to enter the DAG pool, which are subsequently converted to TAG by DGAT and PDAT. Below we describe some variants of these enzymes that are specialized for incorporating different types of specialty FAs into TAGs.

LPATs with substrate specificities for specialty FAs

Divergent LPATs from specialty FA-accumulating organisms have evolved specialized substrate specificities for incorporating specialty FAs into TAGs (Table 2). For instance, LPAT variants from organisms naturally accumulating MCFAs (e.g., *C. nucifera*, *C. viscosissima*, and *C. pulcherrima*) preferentially incorporate MCFAs to the *sn*-2 position of LPA, and when combined with FATBs, enabled efficient deposition of MCFAs at the *sn*-2 position of TAG and further increased total MCFA contents (Knutzon et al., 1999; Kim et al., 2015a; Kim et al., 2015b). Notably, expression of *CnLPAT*, *CvLPAT2*, and *CpuLPAT2a* resulted in the deposition of capric acid (C10:0) at the TAG *sn*-2 position, whereas *CpuLPATB* expression led to accumulation of myristic acid (C14:0) instead of capric acid (C10:0) at the *sn*-2 position of TAG, suggesting distinct substrate specificities of divergent forms of LPATs for different MCFAs

(Knutzon et al., 1999; Kim et al., 2015a; Kim et al., 2015b). RcLPAT2 isolated from castor and VfLPAT2 from tung tree producing conjugated FAs improved the accumulation of hydroxy FAs and conjugated FAs, respectively, in Arabidopsis seeds (Shockey et al., 2019). Co-expression of the *LPAT* from epoxy FA-rich *V. galamensis* with *VgEPX* increased the level of epoxy FAs from 8.7% (*VgEPX* alone) to 16.7% of total lipids in *N. benthamiana* leaves (Sun et al., 2022). For CPA engineering, co-expression of *SfLPAT2* from CPA-enriched *S. foetida* with *EcCPS* in the *fad2/fae1* mutant resulted in the accumulation of CPA at both *sn*-1 and *sn*-2 positions of PC and further increased CPA contents to 35% and 18% of seed oils in Arabidopsis and Camelina, respectively (Yu et al., 2014; Yu et al., 2018).

LPCAT, PDCT and PLC variants channeling specialty FA into PC and DAG

Given the crucial role of PC in acyl editing and TAG biosynthesis, specialized LPCAT, the enzyme catalyzing the conversion of lysophosphatidylcholine (LPC) and acyl-CoA to PC, may contribute to the incorporation of specialty FAs into TAGs. Indeed, the specialized LPCAT from *V. galamensis*, an oleaginous plant containing high levels of epoxy FAs in its seed oil, greatly enhanced the accumulation of epoxy FAs from 8.7% to as much as 19.4% of total lipids when co-expressed with *VgEPX* in *N. benthamiana* leaves (Sun et al., 2022). Studies of transgenic Arabidopsis and Camelina engineered to produce CPA and hydroxy FAs revealed that PCs containing these specialty FAs were not efficiently converted to DAGs and TAGs, identifying bottlenecks for the accumulation of specialty FAs (Bates and Browse, 2011; Yu et al., 2018). To address these bottlenecks, specialized enzymes that convert CPA-containing or hydroxy-containing PCs to DAGs including PDCT and PLC were used to further enhance the production of specialty FAs (Figure 1; Table 2). In efforts to engineer hydroxy FAs in model and crop plants, a PDCT (RcPDCT) and a PLC (RcPLCL1) were isolated from castor and tested in transgenic plants expressing *RcFAH12*. It was shown that both RcPDCT and RcPLCL1 could enrich hydroxy FA in DAG and TAG and increase total hydroxy FA contents from 10-15% to approximately 20% of seed oil (Hu et al., 2012; Aryal and Lu, 2018). Co-expression of *LcPDCT* from CPA-enriched *L. chinensis* with *EcCPS* enhanced the deposition of CPA in DAG and TAG and led to a 50% increase of CPA in seed oil compared with that of plants expressing *EcCPS* alone (Yu et al., 2019).

Specialized DGATs and PDATs for producing TAGs containing specialty FAs

Several studies have reported that specialized DGATs and PDATs with high specificities for specialty FAs are necessary for addressing bottlenecks for the accumulation of specialty FAs in target plants (Park et al., 2021; Lunn et al., 2022). In one of such study, a DGAT from *C. pulcherrima*, namely CpuDGAT1 was identified, which showed a higher enzyme activity toward MCFA substrates relative to typical FAs. Expression of *CpuDGAT1* in Camelina seeds containing MCFA produced by the exogenous CvFATB1 enriched MCFA (capric acid, 10:0) in TAG and increased capric acid content from 8% to 14.5% of seed oil (Iskandarov et al., 2017). Two DGATs from *V. galamensis* were tested in petunia leaves and soybean seeds for their ability to enhance epoxy FA accumulation. When co-expressed with the epoxygenase gene from *Stokesia laevis* (*SlEPX*), both *VgDGAT1* and *VgDGAT2* further increased epoxy FA contents in petunia leaves and soybean seeds, and VgDGAT2 seemed to have a greater impact on epoxy FA accumulation than VgDGAT1 (Li et al., 2010). To engineer conjugated FAs in Arabidopsis, *VfDGAT2*, a DGAT gene isolated from tung tree was co-expressed with *VfFADX*. While no significant increase in eleostearic acid was detected in seeds co-expressing *VfDGAT2* and *VfFADX* relative to that in seeds expressing *VfFADX* alone, introducing *VfDGAT2* into Arabidopsis leaves expressing *VfFADX* resulted in redirection of eleostearic acids from phospholipids to TAGs, an increase in eleostearic acid contents, and mitigation of negative growth effects (Yurchenko et al., 2017; Shockey et al., 2019). For hydroxy FA engineering, specialized DGAT and PDAT isolated from *R. communis* were combined with RcFAH12 individually to produce TAGs with high levels of hydroxy FAs, and both RcDGAT2 and RcPDAT1A enhanced the incorporation of hydroxy FAs into TAG and increased the content of hydroxy FAs to 30% and 27% of seed oil, respectively (Burgal et al., 2008; van Erp et al., 2011; Shockey et al., 2019).

Combinations of TAG-assembly enzymes to enrich specialty FA in TAG

To further enhance the production of specialty TAGs, enzymes involved in different steps of TAG assembly were combined to maximize the incorporation of the specialty FAs into TAGs (Table 2). Combining CvLPAT2 and CpuDGAT1 from MCFA-enriched Cuphea species greatly enriched capric acid (C10:0) accumulation in TAG and increased capric acid content to 23.7% of seed oil, which is higher than that in Camelina expressing these enzymes individually (Iskandarov et al., 2017). For hydroxy FA engineering, RcLPAT2 and RcDGAT2 isolated from castor synergistically increased the level of hydroxy FA to up to 30% of seed oil, but adding RcGPAT9, a specialized GPAT that incorporates hydroxy FAs to the *sn*-1 position of G3P, to this combination did not further boost the accumulation of hydroxy FAs (Shockey et al., 2019). In a similar study, Lunn et al. (2019) successfully enriched tri-hydroxy-TAG, increased hydroxy FA to 34% of seed oil, and restored seed oil content to wild-type levels by co-expressing *RcGPAT9*, *RcLPAT2*, *RcPDAT1A* in an *RcFAH12* transformed Arabidopsis *fae1* mutant (Table 2). Another combination including RcLPCAT, RcPDCT, RcPDAT1-2, RcDGAT2 produced about 31% hydroxy FA in Arabidopsis *fae1* mutant

seeds expressing RcFAH12 and increased both seed size and oil per seed (Park et al., 2022).

Producing acetyl-TAG using specialized DGATs

Acetyl-TAGs (acTAGs) are specialty TAGs with an acetate esterified to the *sn*-3 position in place of a long-chain fatty acid. Oils containing acTAGs exhibit reduced viscosity and therefore have high value in a wide variety of industrial applications such emulsifiers and lubricants. Specialized DGATs (DAcT) responsible for acTAG biosynthesis were isolated from *Euonymus alatus* and *Euonymus fortune*, plants that naturally produce acTAGs in their seeds. Heterologous expression of *EaDAcT* resulted in accumulation of 40% and 52% of acTAG in seeds of Arabidopsis and Camelina, respectively (Durrett et al., 2010). EfDAcT, functioning more efficiently than EaDAcT, produced an average of 72% acTAG in transgenic Camelina seeds (Alkotami et al., 2021). Interestingly, co-expression of *CnLPAT* from MCFA-containing coconut and *EaDAcT* from acTAG-enriched *E. alatus* in camelina plants expressing *UcFATB1* produced acTAGs with MCFAs, which have yet to be found in nature, suggesting a potential synthetic-biology strategy for creating novel lipid structures in plants (Bansal et al., 2018).

Enhancing TAG accumulation by introducing an effective DGAT

DGAT enzymes catalyze the final committed step of TAG biosynthesis, and an efficient DGAT is key to enhancing TAG accumulation in plants (Figure 1). Zienkiewicz et al. (2017) screened six out of 12DGATs from *Nannochloropsis oceanica*, a microalga that produces high amounts of TAGs, and identified DGAT5 (DGTT5) as the most efficient isoform for restoring TAG synthesis in a TAG synthesis-deficient mutant of yeast. Transient expression of *NoDGTT5* in *N. benthamiana* leaves led to a 2-fold increase in TAG, and stable expression of *NoDGTT5* in Arabidopsis increased leaf TAG contents by 6-fold and boosted seed oil content by 50% (Zienkiewicz et al., 2017). In another study of DGATs from microalga, DGAT1 from *Chlorella ellipsoidea* increased the oil content by 8–37% and by 12–18% in seeds of Arabidopsis and *B. napus*, respectively (Guo et al., 2017). In addition, mouse (*Mus musculus*) DGAT2, the predominant DGAT responsible for TAG biosynthesis in mouse, when transiently expressed in *N. benthamiana* leaves, produced over 20-fold more TAG than that of control leaves (Cai et al., 2019). Recently, a study of *Cyperus esculentus*, a unique plant accumulating large amounts of TAG in its underground tubers, revealed that its heterologous expression in *N. tabacum* increased the TAG content to 5.5% of leaf dry weight (DW), which is 7.2-fold and 1.7-fold higher than that in wild-type leaves and leaves expressing *AtDGAT1*, respectively (Gao et al., 2021). Moreover, *CeDGAT2-2* expression resulted in a substantial increase in the proportion of oleic acid in *N. tabacum* leaves (Gao et al., 2021). In another study, heterologous expression of Arabidopsis *DGAT1* reportedly led to a 7-fold increase in TAG contents in *N. tabacum* leaves (Bouvier-Navé et al., 2000). That the DGATs tested in different studies were driven by different promoters, expressed either transiently or stably, and tested in different plant tissues and species, precludes us from making meaningful comparisons for assessing the relative efficacy of DGATs from different sources. Thus, it would be useful to evaluate all promising DGATs under same conditions and in same tissues and target organisms.

Section III. Packaging storage lipids into lipid droplets and reducing degradation

It has been demonstrated in Arabidopsis that FA degradation proceeds *via* a TAG intermediate (Fan et al., 2014). Emerging evidence indicates that proper and efficient packaging of TAGs into LDs is critical for increasing the capacity of lipid accumulation in plant cells, and some lipogenic factors involved in this process have been included in metabolic engineering strategies to enhance lipid production in plants (Table 3). In this section, we describe some attempts to enhance specialty lipid accumulation in plants with the use of LD-related factors and list other LD-related factors that could be engineered for specialty lipid accumulation in plants.

Oleosins (OLE), the predominant LD coat proteins specific to plants, have been used in several studies to engineer LDs for increased lipid accumulation in plant cells. The L-oleosin from sesame (*Sesamum indicum*) and especially its modified version (cysteine- [Cys]-oleosin) have been combined with other lipogenic factors to engineer storage lipids in vegetative tissues of Arabidopsis, *N. tabacum*, *Solanum tuberosum*, and *Sorghum bicolor* (Winichayakul et al., 2013; Vanhercke et al., 2014; Vanhercke et al., 2017; Liu et al., 2017; Vanhercke et al., 2019a). Expression of the castor *RcOLE* in *RcFAH12*-expressing Arabidopsis further increased hydroxy FA from 18% to 22% of seed oil (Lu et al., 2006).

SEIPIN, a key protein that orchestrates the machinery of LD biogenesis at the ER, can promote LD biogenesis and increase TAG contents in plants (Cai et al., 2015). Overexpression of *AtSEIPIN1* in Arabidopsis seeds engineered to synthesize hydroxy FAs increased the hydroxy FA and total lipid contents, representing a potential new target for engineering specialty FAs in plants (Lunn et al., 2018). Interestingly, some LD proteins without apparent homologs in plants still exhibit conserved functional features as part of the LD biogenesis

machinery when ectopically expressed in plants and thus can be used as tools to manipulate LD formation in plants. For instance, ectopic expression of the mouse fat storage-inducing transmembrane protein 2 (FIT2), an ER-localized protein that facilitates the portioning of TAGs from the ER into nascent LDs, in Arabidopsis and *N. benthamiana* led to increased numbers and sizes of LDs and enhanced lipid accumulation in both leaves and seeds (Cai et al., 2017). In another similar study, mouse fat-specific protein 27 (FSP27), a vertebrate-specific protein that mediates LD fusion, was found to promote LD fusion, and enhance the accumulation of LDs and TAGs when expressed in Arabidopsis and *N. benthamiana* (Price et al., 2020). It is an open question whether proteins related to LD formation have evolved specificities for packaging selected specialty TAGs into LDs. Future efforts to elucidate the roles of LD-related proteins in specialty FA accumulation will shed new light on metabolic engineering of desirable lipids in plants.

LD-associated lipases hydrolyze TAGs to release FAs, which are subsequently catabolized *via* β-oxidation in the peroxisomes to produce acetyl-CoA (Eastmond and Graham, 2001). SUGAR DEPENDENT 1 (SDP1) is a primary TAG lipase responsible for TAG degradation in plants (Eastmond, 2006; Kelly et al., 2013b). The suppression of *SDP1* during seed development resulted in increased production of seed oil in Arabidopsis (van Erp et al., 2014), *B. napus* (Kelly et al., 2013a), *Jatropha curcas* (Kim et al., 2014), and soybean (Kanai et al., 2019; Aznar-Moreno et al., 2022). SDP1 from *Physaria fendleri* has been shown to preferentially hydrolyze TAGs containing hydroxy FAs and suppression of its expression increased total FA content by 14-19%, primarily contributing to the significantly increased hydroxy FA (Azeez et al., 2022). Recent studies identified additional proteins involved in the mobilization of LDs in plants including UBX-domain containing protein 10 (PUX10), CELL DIVISION CYCLE 48, (CDC48A), Comparative Gene Identification-58 (CGI58), ATP-binding cassette transporter-like protein (PXA1), and AT-hook motif containing nuclear localized transcriptional repressor (AHL4) (Zolman et al., 2001; James et al., 2010; Park et al., 2013; Deruyffelaere et al., 2018; Kretzschmar et al., 2018; Cai et al., 2020). Future work to tune the expression of these factors may contribute further to enhancing the accumulation of lipids, including specialty FAs, in plants.

Concluding remarks and future perspectives

Extensive efforts and substantial progress have been made in the past two decades to design and test metabolic engineering strategies for producing desirable lipids in plants for bioenergy, industrial, and nutraceutical purposes. These studies have generated a broad array of lipogenic factors for engineering different types of lipids in various plant species. Selecting lipogenic factors that outperform their alternatives when expressed in a target crop is key to optimizing the design of engineering approaches for maximized production of selected lipids. Notably, the optimal lipogenic factors for plant lipid engineering may be sourced outside of the plant kingdom. For instance, the CPS from *E. coli* and the FAH from a fungal pathogen (*C. purpurea*) were shown to be more effective in producing CPA or hydroxy FA in plants than the plant-sourced ones (Meesapyodsuk and Qiu, 2008; Yu et al., 2014). In organisms producing high levels of specialty FAs, besides the enzymes responsible for FA synthesis and modification, other lipogenic factors function in glycerolipid assembly and LD formation may have evolved specialized features to accommodate these specialty FAs by depositing them in TAGs and subsequently packaging them in LDs. Therefore, future efforts to enhance specialty lipid accumulation in agronomic crops may be enhanced by introducing multiple specialized lipogenic factors involved in all key steps in lipid synthesis and packaging.

As our understanding of the structural basis of specialized lipogenic factors increases, future research of metabolic engineering will benefit from designing novel lipogenic factors that can outperform naturally occurring ones or produce novel lipids that have not been previously identified in nature based on sequence comparison, computational protein design or directed evolution. Deployment of new computational tools such as AlphaFold to these efforts will likely enhance their success (Mirdita et al., 2022). The feasibility of the former approach has been validated in several studies. In attempts to generate novel DGAT enzymes with improved efficiencies in TAG production, mutant variants of soybean and hazelnut (*Corylus americana*) DGAT1s produced higher levels of TAGs when expressed in plants compared to the wild-type versions (Roesler et al., 2016; Hatanaka et al., 2022). The structural details of acyl-ACP desaturases guided the generation of a mutant Δ9-acyl-ACP with amino acid substitutions in the substrate binding pocket, which was combined with other lipogenic factors to engineer the specialty ω^7 monounsaturated FAs in seed oil (Cahoon and Shanklin, 2000; Nguyen et al., 2010) (Whittle and Shanklin, 2001). Similarly, expression of the native *M. charantia* FADX in Arabidopsis *fad3/fae1* mutant yielded 10% α-eleostearic acid, while the mutagenized McFADX (G111V) or McFADX (G111V/D115E) resulted in a doubling of conjugated FA accumulation to approximately 20% of seed oil. Like the native McFADX, the mutant McFADX (G111V) produced predominantly α-eleostearic acid and little punicic acid, whereas the McFADX (G111V/D115E) double mutant produced nearly equal amounts of α-eleostearic acid and punicic acid (Rawat et al., 2012). In addition, variants of the castor stearoyl-ACP desaturase (T117R/D280K) generated by site-directed mutation can synthesize a novel FA, *erythro*-9,10-dihydroxystearate, with

vicinal hydroxyl groups at C9 and C10 positions (Whittle et al., 2020). Improved mechanistic understanding will facilitate the development of novel improved lipogenic factors *via* site-directed mutagenesis i.e., rational, structure-based design in combination with computational modeling (Guy et al., 2022) that can be optimized by design-build-test-learn cycles for plant lipid engineering.

Whereas the majority of plant lipid metabolic engineering has focused on seeds, there is growing interest in engineering lipids in plant vegetative tissues because of their high biomass and high capacity for FA synthesis. Most of the FA flux in plant vegetative tissues is for phospholipids to support membrane synthesis, while TAGs serve as an intermediate for FA degradation and are present only at a minimal level in vegetative tissues (Fan et al., 2014). A variety of lipogenic factors and combinations thereof, have been evaluated for their efficacies in enhancing storage lipid accumulation in vegetative tissues of a small number of plant species (Vanhercke et al., 2019b). So far, the most successful approach, characterized as the "push, pull, and protect" strategy, include 1) seed-specific transcription factors such as WRINKLED1 and LEAFY COTYLEDON2 (LEC2) to push the carbon flux toward FA synthesis, 2) acyltransferases such as DGAT and PDAT to pull FAs into the TAG pool, and 3) LD proteins such as oleosin to package TAGs into LDs and protect them from degradation (Vanhercke et al., 2014; Zale et al., 2016; Vanhercke et al., 2017; Alameldin et al., 2017; Liu et al., 2017; Vanhercke et al., 2019a). A major challenge for enhancing TAG accumulation in non-seed tissues is the impairment of growth associated with TAG accumulation, which may result from the accumulation of cytotoxic free FAs, toxic effects of expression of seed-specific transcription factors, and/or the enlarged TAG pool redirecting carbon flux away from other metabolic pathways (Yang et al., 2015; Zale et al., 2016; Vanhercke et al., 2019a; Mitchell et al., 2020). Future efforts to develop improved strategies for mitigated growth impairment and further enhancement of vegetative TAG production will focus on the identification of alternative lipogenic factors that can more efficiently incorporate FAs to TAGs and have reduced negative impacts on plant growth. Additional promising approaches include restricting the expression of lipogenic factors to certain tissues or growth stages using inducible or tissue-specific promoters (Andrianov et al., 2010; Kim et al., 2015c; Liang et al., 2022) , or the expression of factors such as purple acid phosphatase2 (Cai et al., 2022). Despite the challenges, vegetative biomass represents a sustainable and economical platform for lipid accumulation and the success in engineering TAG accumulation therein will facilitate increased yields per unit land area of high-value lipids containing specialty FAs in vegetative tissues by introducing additional specialized lipogenic factors.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

Author contributions

X-HY, YC, and JS conceived the study; YC and X-HY drafted the manuscript; JS revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Funding support was provided by the DOE Center for Advanced Bioenergy and Bioproducts Innovation (the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0018420) and Genomic Science Program (the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, grant no. DE-SC0021369). JS was funded in part by the Physical Biosciences Program, within the US Department of Energy (DOE), Division of Chemical Sciences, Geosciences and Biosciences (grant KC0304000).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Author disclaimer

Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the U.S. Department of Energy.

References

Abbadi, A., Domergue, F., Bauer, J., Napier, J. A., Welti, R., Zähringer, U., et al. (2004). Biosynthesis of very-long-chain polyunsaturated fatty acids in transgenic oilseeds: constraints on their accumulation. *Plant Cell* 16, 2734–2748.

Alameldin, H., Izadi-Darbandi, A., Smith, S. A., Balan, V., Jones, A. D., and Sticklen, M. (2017). Production of seed-like storage lipids and increase in oil bodies in corn (Maize; zea mays l.) vegetative biomass. *Ind. Crops Products.* 108, 526–534.

Alkotami, L., Kornacki, C., Campbell, S., McIntosh, G., Wilson, C., Tran, T. N. T., et al. (2021). Expression of a high-activity diacylglycerol acetyltransferase results in enhanced synthesis of acetyl-TAG in camelina seed oil. *Plant J.* 106, 953–964.

Andrianov, V., Borisjuk, N., Pogrebnyak, N., Brinker, A., Dixon, J., Spitsin, S., et al. (2010). Tobacco as a production platform for biofuel: overexpression of arabidopsis DGAT and LEC2 genes increases accumulation and shifts the composition of lipids in green biomass. *Plant Biotechnol. J.* 8, 277–287.

Aryal, N., and Lu, C. (2018). A phospholipase c-like protein from ricinus communis increases hydroxy fatty acids accumulation in transgenic seeds of camelina sativa. *Front. Plant Sci.* 9, 1576.

Azeez, A., Parchuri, P., and Bates, P. D. (2022). Suppression of physaria fendleri SDP1 increased seed oil and hydroxy fatty acid content while maintaining oil biosynthesis through triacylglycerol remodeling. *Front. Plant Sci.* 13, 931310.

Aznar-Moreno, J. A., Mukherjee, T., Morley, S. A., Duressa, D., Kambhampati, S., Chu, K. L., et al. (2022). Suppression of SDP1 improves soybean seed composition by increasing oil and reducing undigestible oligosaccharides. *Front. Plant Sci.* 13, 863254.

Bafor, M., Smith, M. A., Jonsson, L., Stobart, K., and Stymne, S. (1993). Biosynthesis of vernoleate (cis-12-epoxyoctadeca-cis-9-enoate) in microsomal preparations from developing endosperm of euphorbia lagascae. *Arch. Biochem. Biophys.* 303, 145–151.

Bansal, S., Kim, H. J., Na, G., Hamilton, M. E., Cahoon, E. B., Lu, C., et al. (2018). Towards the synthetic design of camelina oil enriched in tailored acetyl-triacylglycerols with medium-chain fatty acids. *J. Exp. Bot.* 69, 4395–4402.

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

Bao, X., Thelen, J. J., Bonaventure, G., and Ohlrogge, J. B. (2003). Characterization of cyclopropane fatty-acid synthase from sterculia foetida. *J. Biol. Chem.* 278, 12846–12853.

Bates, P. D., and Browse, J. (2011). The pathway of triacylglycerol synthesis through phosphatidylcholine in arabidopsis produces a bottleneck for the accumulation of unusual fatty acids in transgenic seeds. *Plant J.* 68, 387–399.

Bondaruk, M., Johnson, S., Degafu, A., Boora, P., Bilodeau, P., Morris, J., et al. (2007). Expression of a cDNA encoding palmitoyl-acyl carrier protein desaturase from cat's claw (Doxantha unguis-cati l.) in arabidopsis thaliana and brassica napus leads to accumulation of unusual unsaturated fatty acids and increased stearic acid content in the seed oil. *Plant Breed.* 126, 186–194.

Bouvier-Navé, P., Benveniste, P., Oelkers, P., Sturley, S. L., and Schaller, H. (2000). Expression in yeast and tobacco of plant cDNAs encoding acyl CoA: diacylglycerol acyltransferase. *Eur. J. Biochem.* 267, 85–96.

Broadwater, J. A., Whittle, E., and Shanklin, J. (2002). Desaturation and hydroxylation. residues 148 and 324 of arabidopsis FAD2 in addition to substrate chain length exert a major influence in partitioning of catalytic specificity. *J. Biol. Chem.* 277, 15613–15620.

Broun, P., Shanklin, J., Whittle, E., and Somerville, C. (1998). Catalytic plasticity of fatty acid modification enzymes underlying chemical diversity of plant lipids. *Science* 282, 1315–1317.

Broun, P., and Somerville, C. (1997). Accumulation of ricinoleic lesquerolic and densipolic acids in seeds of transgenic arabidopsis plants that express a fatty acyl hydroxylase cDNA from castor bean. *Plant Physiol.* 113, 933–942.

Burgal, J., Shockey, J., Lu, C., Dyer, J., Larson, T., Graham, I., et al. (2008). Metabolic engineering of hydroxy fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. *Plant Biotechnol. J.* 6, 819–831.

Cahoon, E. B., Carlson, T. J., Ripp, K. G., Schweiger, B. J., Cook, G. A., Hall, S. E., et al. (1999). Biosynthetic origin of conjugated double bonds: production of fatty acid components of high-value drying oils in transgenic soybean embryos. *Proc. Natl. Acad. Sci. U.S.A.* 96, 12935–12940.

Cahoon, E. B., Coughlan, S. J., and Shanklin, J. (1997). Characterization of a structurally and functionally diverged acyl-acyl carrier protein desaturase from milkweed seed. *Plant Mol. Biol.* 33, 1105–1110.

Cahoon, E. B., Dietrich, C. R., Meyer, K., Damude, H. G., Dyer, J. M., and Kinney, A. J. (2006). Conjugated fatty acids accumulate to high levels in phospholipids of metabolically engineered soybean and arabidopsis seeds. *Phytochemistry* 67, 1166–1176.

Cahoon, E. B., Ripp, K. G., Hall, S. E., and McGonigle, B. (2002). Transgenic production of epoxy fatty acids by expression of a cytochrome P450 enzyme from euphorbia lagascae seed. *Plant Physiol.* 128, 615–624.

Cahoon, E. B., Shah, S., Shanklin, J., and Browse, J. (1998). A determinant of substrate specificity predicted from the acyl-acyl carrier protein desaturase of developing cat's claw seed. *Plant Physiol.* 117, 593–598.

Cahoon, E. B., and Shanklin, J. (2000). Substrate-dependent mutant complementation to select fatty acid desaturase variants for metabolic engineering of plant seed oils. *Proc. Natl. Acad. Sci. U.S.A.* 97, 12350–12355.

Cai, Y., Goodman, J. M., Pyc, M., Mullen, R. T., Dyer, J. M., and Chapman, K. D. (2015). Arabidopsis SEIPIN proteins modulate triacylglycerol accumulation and influence lipid droplet proliferation. *Plant Cell* 27, 2616–2636.

Cai, G., Kim, S. C., Li, J., Zhou, Y., and Wang, X. (2020). Transcriptional regulation of lipid catabolism during seedling establishment. *Mol. Plant* 13, 984–1000.

Cai, Y., McClinchie, E., Price, A., Nguyen, T. N., Gidda, S. K., Watt, S. C., et al. (2017). Mouse fat storage-inducing transmembrane protein 2 (FIT2) promotes lipid droplet accumulation in plants. *Plant Biotechnol. J.* 15, 824–836.

Cai, Y., Whitehead, P., Chappell, J., and Chapman, K. D. (2019). Mouse lipogenic proteins promote the co-accumulation of triacylglycerols and sesquiterpenes in plant cells. *Planta* 250, 79–94.

Cai, Y., Zhai, Z., Keereetaweep, J., Liu, H., Shi, H., Schwender, J., et al (2022). Purple acid phosphatase 2 increases fatty acid synthesis and offsets yield drag associated with triacylglycerol hyperaccumulation in vegetative tissues. *New Phytologist (CABBI and BES).* doi: 10.1111/nph.18392

Carlsson, A. S., Yilmaz, J. L., Green, A. G., Stymne, S., and Hofvander, P. (2011). Replacing fossil oil with fresh oil - with what and for what? *Eur. J. Lipid Sci. Technol.* 113, 812–831.

Chapman, K. D., Dyer, J. M., and Mullen, R. T. (2012). Biogenesis and functions of lipid droplets in plants: Thematic review series: Lipid droplet synthesis and metabolism: from yeast to man. *J. Lipid Res.* 53, 215–226.

Chen, G. Q., Johnson, K., Nazarenus, T. J., Ponciano, G., Morales, E., and Cahoon, E. B. (2021). Genetic engineering of lesquerella with increased ricinoleic acid content in seed oil. *Plants (Basel).* 10 (6), 1093.

Dahlqvist, A., Stahl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., et al. (2000). Phospholipid:diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6487–6492.

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

Dehesh, K., Jones, A., Knutzon, D. S., and Voelker, T. A. (1996). Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by overexpression of ch FatB2 a thioesterase cDNA from cuphea hookeriana. *Plant J.* 9, 167–172.

Deruyffelaere, C., Purkrtova, Z., Bouchez, I., Collet, B., Cacas, J. L., Chardot, T., et al. (2018). PUX10 is a CDC48A adaptor protein that regulates the extraction of ubiquitinated oleosins from seed lipid droplets in arabidopsis. *Plant Cell* 30, 2116–2136.

Durrett, T. P., McClosky, D. D., Tumaney, A. W., Elzinga, D. A., Ohlrogge, J., and Pollard, M. (2010). A distinct DGAT with sn-3 acetyltransferase activity that synthesizes unusual reduced-viscosity oils in euonymus and transgenic seeds. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9464–9469.

Dyer, J. M., Chapital, D. C., Kuan, J. C., Mullen, R. T., Turner, C., McKeon, T. A., et al. (2002). Molecular analysis of a bifunctional fatty acid conjugase/desaturase from tung. implications for the evolution of plant fatty acid diversity. *Plant Physiol.* 130, 2027–2038.

Dyer, J. M., Stymne, S., Green, A. G., and Carlsson, A. S. (2008). High-value oils from plants. *Plant J.* 54, 640–655.

Eastmond, P. J. (2006). SUGAR-DEPENDENT1 encodes a patatin domain triacylglycerol lipase that initiates storage oil breakdown in germinating arabidopsis seeds. *Plant Cell* 18, 665–675.

Eastmond, P. J., and Graham, I. A. (2001). Re-examining the role of the glyoxylate cycle in oilseeds. *Trends Plant Sci.* 6, 72–78.

Eccleston, V. S., Cranmer, A. M., Voelker, T. A., and Ohlrogge, J. B. (1996). Medium-chain fatty acid biosynthesis and utilization in brassica napus plants expressing lauroyl-acyl carrier protein thioesterase. *Planta* 198, 46–53.

Fan, J., Yan, C., Roston, R., Shanklin, J., and Xu, C. (2014). Arabidopsis lipins PDAT1 acyltransferase and SDP1 triacylglycerol lipase synergistically direct fatty acids toward β-oxidation thereby maintaining membrane lipid homeostasis. *Plant Cell* 26, 4119–4134.

Gao, Y., Sun, Y., Gao, H., Chen, Y., Wang, X., Xue, J., et al. (2021). Ectopic overexpression of a type-II DGAT (CeDGAT2-2) derived from oil-rich tuber of cyperus esculentus enhances accumulation of oil and oleic acid in tobacco leaves. *Biotechnol. Biofuels* 14, 76.

Guo, X., Fan, C., Chen, Y., Wang, J., Yin, W., Wang, R. R., et al. (2017). Identification and characterization of an efficient acyl-CoA: diacylglycerol acyltransferase 1 (DGAT1) gene from the microalga chlorella ellipsoidea. *BMC Plant Biol.* 17, 48.

Guy, J. E., Cai, Y., Baer, M. D., Whittle, E., Chai, J., Yu, X. H., et al. (2022). Regioselectivity mechanism of the thunbergia alata Δ6-16:0-acyl carrier protein desaturase. *Plant Physiol.* 188, 1537–1549.

Han, L., Silvestre, S., Sayanova, O., Haslam, R. P., and Napier, J. A. (2022). Using field evaluation and systematic iteration to rationalize the accumulation of omega-3 long-chain polyunsaturated fatty acids in transgenic camelina sativa. *Plant Biotechnol. J* 20 (9), 1833–1852.

Han, L., Usher, S., Sandgrind, S., Hassall, K., Sayanova, O., Michaelson, L. V., et al. (2020). High level accumulation of EPA and DHA in field-grown transgenic camelina - a multi-territory evaluation of TAG accumulation and heterogeneity. *Plant Biotechnol. J.* 18, 2280–2291.

Hatanaka, T., Shimizu, R., and Hildebrand, D. (2004). Expression of a stokesia laevis epoxygenase gene. *Phytochemistry* 65, 2189–2196.

Hatanaka, T., Tomita, Y., Matsuoka, D., Sasayama, D., Fukayama, H., Azuma, T., et al. (2022). Different acyl-CoA:diacylglycerol acyltransferases vary widely in function and a targeted amino acid substitution enhances oil accumulation. *J. Exp. Bot.* 73, 3030–3043.

Horn, P. J., Liu, J., Cocuron, J. C., McGlew, K., Thrower, N. A., Larson, M., et al. (2016). Identification of multiple lipid genes with modifications in expression and sequence associated with the evolution of hydroxy fatty acid accumulation in physaria fendleri. *Plant J.* 86, 322–348.

Hu, Z., Ren, Z., and Lu, C. (2012). The phosphatidylcholine diacylglycerol cholinephosphotransferase is required for efficient hydroxy fatty acid accumulation in transgenic arabidopsis. *Plant Physiol.* 158, 1944–1954.

Iskandarov, U., Silva, J. E., Kim, H. J., Andersson, M., Cahoon, R. E., Mockaitis, K., et al. (2017). A specialized diacylglycerol acyltransferase contributes to the extreme medium-chain fatty acid content of. *Plant Physiol.* 174, 97–109.

James, C. N., Horn, P. J., Case, C. R., Gidda, S. K., Zhang, D., Mullen, R. T., et al. (2010). Disruption of the arabidopsis CGI-58 homologue produces chanarin-dorfman-like lipid droplet accumulation in plants. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17833–17838.

Kalinger, R. S., Pulsifer, I. P., Hepworth, S. R., and Rowland, O. (2020). Fatty acyl synthetases and thioesterases in plant lipid metabolism: Diverse functions and biotechnological applications. *Lipids* 55, 435–455.

Kalinger, R. S., Williams, D., Ahmadi Pirshahid, A., Pulsifer, I. P., and Rowland, O. (2021). Production of C6-C14 medium-chain fatty acids in seeds and leaves *via* overexpression of single hotdog-fold acyl-lipid thioesterases. *Lipids* 56, 327–344.

Kallio, P., Pásztor, A., Akhtar, M. K., and Jones, P. R. (2014). Renewable jet fuel. *Curr. Opin. Biotechnol.* 26, 50–55.

Kanai, M., Yamada, T., Hayashi, M., Mano, S., and Nishimura, M. (2019). Soybean (Glycine max l.) triacylglycerol lipase GmSDP1 regulates the quality and quantity of seed oil. *Sci. Rep.* 9, 8924.

Keereetaweep, J., Liu, H., Zhai, Z., and Shanklin, J. (2018). Biotin attachment domain-containing proteins irreversibly inhibit acetyl CoA carboxylase. *Plant Physiol.* 177, 208–215.

Kelly, A. A., Shaw, E., Powers, S. J., Kurup, S., and Eastmond, P. J. (2013a). Suppression of the SUGAR-DEPENDENT1 triacylglycerol lipase family during seed development enhances oil yield in oilseed rape (Brassica napus l.). *Plant Biotechnol. J.* 11, 355–361.

Kelly, A. A., van Erp, H., Quettier, A. L., Shaw, E., Menard, G., Kurup, S., et al. (2013b). The sugar-dependent1 lipase limits triacylglycerol accumulation in vegetative tissues of arabidopsis. *Plant Physiol.* 162, 1282–1289.

Kim, H. J., Silva, J. E., Iskandarov, U., Andersson, M., Cahoon, R. E., Mockaitis, K., et al. (2015a). Structurally divergent lysophosphatidic acid acyltransferases with high selectivity for saturated medium chain fatty acids from cuphea seeds. *Plant J.* 84, 1021–1033.

Kim, H. J., Silva, J. E., Vu, H. S., Mockaitis, K., Nam, J. W., and Cahoon, E. B. (2015b). Toward production of jet fuel functionality in oilseeds: identification of FatB acyl-acyl carrier protein thioesterases and evaluation of combinatorial expression strategies in camelina seeds. *J. Exp. Bot.* 66, 4251–4265.

Kim, H. U., Lee, K. R., Jung, S. J., Shin, H. A., Go, Y. S., Suh, M. C., et al. (2015c). Senescence-inducible LEC2 enhances triacylglycerol accumulation in leaves without negatively affecting plant growth. *Plant Biotechnol. J.* 13, 1346–1359.

Kim, M. J., Yang, S. W., Mao, H. Z., Veena, S. P., Yin, J. L., and Chua, N. H. (2014). Gene silencing of sugar-dependent 1 (JcSDP1) encoding a patatin-domain triacylglycerol lipase enhances seed oil accumulation in jatropha curcas. *Biotechnol. Biofuels* 7, 36.

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

Kretzschmar, F. K., Mengel, L. A., Müller, A. O., Schmitt, K., Blersch, K. F., Valerius, O., et al. (2018). PUX10 is a lipid droplet-localized scaffold protein that interacts with CELL DIVISION CYCLE48 and is involved in the degradation of lipid droplet proteins. *Plant Cell* 30, 2137–2160.

Lee, J. S., Takai, J., Takahasi, K., Endo, Y., Fujimoto, K., Koike, S., et al. (2002). Effect of dietary tung oil on the growth and lipid metabolism of laying hens. *J. Nutr. Sci. Vitaminol. (Tokyo).* 48, 142–148.

Liang, Y., Yu, X-H., Anaokar, S., Shi, H., Dahl, W.S., Cai, Y., et al (2002). Engineering triacylglycerol accumulation in duckweed (*Lemna japonica*). *Plant Biotechnology Journal Plant Biotechnol J.* doi: 10.1111/pbi.13943

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., et al. (2013). Acyl-lipid metabolism. *Arabidopsis. Book.* 11, e0161.

Liu, Q., Guo, Q., Akbar, S., Zhi, Y., El Tahchy, A., Mitchell, M., et al. (2017). Genetic enhancement of oil content in potato tuber (Solanum tuberosum l.) through an integrated metabolic engineering strategy. *Plant Biotechnol. J.* 15, 56–67.

Liu, L., Hammond, E. G., and Nikolau, B. J. (1998). *In vivo* studies of the biosynthesis of vernolic acid in the seed of vernonia galamensis. *Lipids* 33, 1217–1221.

Li, R., Yu, K., Hatanaka, T., and Hildebrand, D. F. (2010). Vernonia DGATs increase accumulation of epoxy fatty acids in oil. *Plant Biotechnol. J.* 8, 184–195.

Lu, C., Fulda, M., Wallis, J. G., and Browse, J. (2006). A high-throughput screen for genes from castor that boost hydroxy fatty acid accumulation in seed oils of transgenic arabidopsis. *Plant J.* 45, 847–856.

Lunn, D., Wallis, J. G., and Browse, J. (2018). Overexpression of Seipin1 increases oil in hydroxy fatty acid-accumulating seeds. *Plant Cell Physiol.* 59, 205–214.

Lunn, D., Wallis, J. G., and Browse, J. (2019). Tri-Hydroxy-Triacylglycerol is efficiently produced by position-specific castor acyltransferases. *Plant Physiol.* 179, 1050–1063.

Lunn, D., Wallis, J. G., and Browse, J. (2022). A multigene approach secures hydroxy fatty acid production in arabidopsis. *J. Exp. Bot.* 73, 2875–2888.

Lu, C., Xin, Z., Ren, Z., Miquel, M., and Browse, J. (2009). An enzyme regulating triacylglycerol composition is encoded by the ROD1 gene of arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 106, 18837–18842.

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

Meesapyodsuk, D., and Qiu, X. (2008). An oleate hydroxylase from the fungus claviceps purpurea: cloning functional analysis and expression in arabidopsis. *Plant Physiol.* 147, 1325–1333.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682.

Mitchell, M. C., Pritchard, J., Okada, S., Zhang, J., Venables, I., Vanhercke, T., et al. (2020). Increasing growth and yield by altering carbon metabolism in a transgenic leaf oil crop. *Plant Biotechnol. J* 18 (10), 2042–2052.

Napier, J. A., Haslam, R. P., Tsalavouta, M., and Sayanova, O. (2019). The challenges of delivering genetically modified crops with nutritional enhancement traits. *Nat. Plants* 5, 563–567.

Nguyen, H. T., Mishra, G., Whittle, E., Pidkowich, M. S., Bevan, S. A., Merlo, A. O., et al. (2010). Metabolic engineering of seeds can achieve levels of omega-7 fatty acids comparable with the highest levels found in natural plant sources. *Plant Physiol.* 154, 1897–1904.

Nguyen, H. T., Park, H., Koster, K. L., Cahoon, R. E., Shanklin, J., Clemente, T. E., et al. (2015). Redirection of metabolic flux for high levels of omega-7 monounsaturated fatty acid accumulation in camelina seeds. *Plant Biotechnol. J.* 13, 38–50.

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

Ohlrogge, J., and Chapman, K. (2011). The seeds of green energy: expanding the contribution of plant oils as biofuels. *Biochemist.* 33, 34–38.

Okada, S., Taylor, M., Zhou, X. R., Naim, F., Marshall, D., Blanksby, S. J., et al. (2020). Producing cyclopropane fatty acid in plant leafy biomass. *Front. Plant Sci.* 11, 30.

Park, S., Gidda, S. K., James, C. N., Horn, P. J., Khuu, N., Seay, D. C., et al. (2013). The α/β hydrolase CGI-58 and peroxisomal transport protein PXA1 coregulate lipid homeostasis and signaling in arabidopsis. *Plant Cell* 25, 1726–1739.

Park, M. E., Lee, K. R., Chen, G. Q., and Kim, H. U. (2022). Enhanced production of hydroxy fatty acids in arabidopsis seed through modification of multiple gene expression. *Biotechnol. Biofuels Bioprod.* 15, 66.

Park, K., Sanjaya, S. A., Quach, T., and Cahoon, E. B. (2021). Toward sustainable production of value-added bioenergy and industrial oils in oilseed and biomass feedstocks. *GCB. Bioenergy* 13, 1610–1623.

Petrie, J. R., Shrestha, P., Belide, S., Kennedy, Y., Lester, G., Liu, Q., et al. (2014). Metabolic engineering camelina sativa with fish oil-like levels of DHA. *PloS One* 9, e85061.

Petrie, J. R., Shrestha, P., Liu, Q., Mansour, M. P., Wood, C. C., Zhou, X. R., et al. (2010). Rapid expression of transgenes driven by seed-specific constructs in leaf tissue: DHA production. *Plant Methods* 6, 8.

Petrie, J. R., Shrestha, P., Zhou, X. R., Mansour, M. P., Liu, Q., Belide, S., et al. (2012). Metabolic engineering plant seeds with fish oil-like levels of DHA. *PloS One* 7, e49165.

Price, A. M., Doner, N. M., Gidda, S. K., Jambunathan, S., James, C. N., Schami, A., et al. (2020). Mouse fat-specific protein 27 (FSP27) expressed in plant cells localizes to lipid droplets and promotes lipid droplet accumulation and fusion. *Biochimie* 169, 41–53.

Pyc, M., Cai, Y., Greer, M. S., Yurchenko, O., Chapman, K. D., Dyer, J. M., et al. (2017). Turning over a new leaf in lipid droplet biology. *Trends Plant Sci.* 22, 596–609.

Qi, B., Fraser, T., Mugford, S., Dobson, G., Sayanova, O., Butler, J., et al. (2004). Production of very long chain polyunsaturated omega-3 and omega-6 fatty acids in plants. *Nat. Biotechnol.* 22, 739–745.

Rawat, R., Yu, X. H., Sweet, M., and Shanklin, J. (2012). Conjugated fatty acid synthesis: residues 111 and 115 influence product partitioning of momordica charantia conjugase. *J. Biol. Chem.* 287, 16230–16237.

Roesler, K., Shen, B., Bermudez, E., Li, C., Hunt, J., Damude, H. G., et al. (2016). An improved variant of soybean type 1 diacylglycerol acyltransferase increases the oil content and decreases the soluble carbohydrate content of soybeans. *Plant Physiol.* 171, 878–893.

Salie, M. J., Zhang, N., Lancikova, V., Xu, D., and Thelen, J. J. (2016). A family of negative regulators targets the committed step of de Novo fatty acid biosynthesis. *Plant Cell* 28, 2312–2325.

Shanklin, J., and Cahoon, E. B. (1998). DESATURATION AND RELATED MODIFICATIONS OF FATTY ACIDS1. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 49, 611–641.

Shockey, J., Lager, I., Stymne, S., Kotapati, H. K., Sheffield, J., Mason, C., et al. (2019). Specialized lysophosphatidic acid acyltransferases contribute to unusual fatty acid accumulation in exotic euphorbiaceae seed oils. *Planta* 249, 1285–1299.

Singh, R., Arora, A., and Singh, V. (2021). Biodiesel from oil produced in vegetative tissues of biomass - a review. *Bioresour. Technol.* 326, 124772.

Singh, S., Thomaeus, S., Lee, M., Stymne, S., and Green, A. (2001). Transgenic expression of a delta 12-epoxygenase gene in arabidopsis seeds inhibits accumulation of linoleic acid. *Planta* 212, 872–879.

Smith, M. A., Moon, H., Chowrira, G., and Kunst, L. (2003). Heterologous expression of a fatty acid hydroxylase gene in developing seeds of arabidopsis thaliana. *Planta* 217, 507–516.

Sun, Y., Liu, B., Xue, J., Wang, X., Cui, H., Li, R., et al. (2022). Critical metabolic pathways and genes cooperate for epoxy fatty acid-enriched oil production in developing seeds of vernonia galamensis an industrial oleaginous plant. *Biotechnol. Biofuels Bioprod.* 15, 21.

Tjellström, H., Strawsine, M., Silva, J., Cahoon, E. B., and Ohlrogge, J. B. (2013). Disruption of plastid acyl:acyl carrier protein synthetases increases medium chain fatty acid accumulation in seeds of transgenic arabidopsis. *FEBS Lett.* 587, 936–942.

Tocher, D. R., Betancor, M. B., Sprague, M., Olsen, R. E., and Napier, J. A. (2019). Omega-3 long-chain polyunsaturated fatty acids EPA and DHA: Bridging the gap between supply and demand. *Nutrients* 11 (1), 89.

van de Loo, F. J., Broun, P., Turner, S., and Somerville, C. (1995). An oleate 12-hydroxylase from ricinus communis l. *is. Fatty acyl. desaturase. homolog. Proc. Natl. Acad. Sci. U.S.A.* 92, 6743–6747.

van Erp, H., Bates, P. D., Burgal, J., Shockey, J., and Browse, J. (2011). Castor phospholipid:diacylglycerol acyltransferase facilitates efficient metabolism of hydroxy fatty acids in transgenic arabidopsis. *Plant Physiol.* 155, 683–693.

van Erp, H., Kelly, A. A., Menard, G., and Eastmond, P. J. (2014). Multigene engineering of triacylglycerol metabolism boosts seed oil content in arabidopsis. *Plant Physiol.* 165, 30–36.

Vanhercke, T., Belide, S., Taylor, M. C., El Tahchy, A., Okada, S., Rolland, V., et al. (2019a). Up-regulation of lipid biosynthesis increases the oil content in leaves of sorghum bicolor. *Plant Biotechnol. J.* 17, 220–232.

Vanhercke, T., Divi, U. K., El Tahchy, A., Liu, Q., Mitchell, M., Taylor, M. C., et al. (2017). Step changes in leaf oil accumulation *via* iterative metabolic engineering. *Metab. Eng.* 39, 237–246.

Vanhercke, T., Dyer, J. M., Mullen, R. T., Kilaru, A., Rahman, M. M., Petrie, J. R., et al. (2019b). Metabolic engineering for enhanced oil in biomass. *Prog. Lipid Res.* 74, 103–129.

Vanhercke, T., El Tahchy, A., Liu, Q., Zhou, X. R., Shrestha, P., Divi, U. K., et al. (2014). Metabolic engineering of biomass for high energy density: oilseed-like triacylglycerol yields from plant leaves. *Plant Biotechnol. J.* 12, 231–239.

Voelker, T. A., Hayes, T. R., Cranmer, A. M., Turner, J. C., and Davies, H. M. (1996). Genetic engineering of a quantitative trait: metabolic and genetic parameters influencing the accumulation of laurate in rapeseed. *Plant J.* 9, 229–241.

Wang, X. (2001). PLANT PHOSPHOLIPASES. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 52, 211–231.

Whittle, E. J., Cai, Y., Keereetaweep, J., Chai, J., Buist, P. H., and Shanklin, J. (2020). Castor stearoyl-ACP desaturase can synthesize a vicinal diol by dioxygenase chemistry. *Plant Physiol.* 182, 730–738.

Whittle, E., and Shanklin, J. (2001). Engineering delta 9-16:0-acyl carrier protein (ACP) desaturase specificity based on combinatorial saturation mutagenesis and logical redesign of the castor delta 9-18:0-ACP desaturase. *J. Biol. Chem.* 276, 21500–21505.

Winichayakul, S., Scott, R. W., Roldan, M., Hatier, J. H., Livingston, S., Cookson, R., et al. (2013). *In vivo* packaging of triacylglycerols enhances arabidopsis leaf biomass and energy density. *Plant Physiol.* 162, 626–639.

Wood, C. C., Petrie, J. R., Shrestha, P., Mansour, M. P., Nichols, P. D., Green, A. G., et al. (2009). A leaf-based assay using interchangeable design principles to rapidly assemble multistep recombinant pathways. *Plant Biotechnol. J.* 7, 914–924.

Wu, G., Truksa, M., Datla, N., Vrinten, P., Bauer, J., Zank, T., et al. (2005). Stepwise engineering to produce high yields of very long-chain polyunsaturated fatty acids in plants. *Nat. Biotechnol.* 23, 1013–1017.

Xu, Y., Mietkiewska, E., Shah, S., Weselake, R. J., and Chen, G. (2020). Punicic acid production in brassica napus. *Metab. Eng.* 62, 20–29.

Xu, C., and Shanklin, J. (2016). Triacylglycerol metabolism function and accumulation in plant vegetative tissues. *Annu. Rev. Plant Biol.* 67, 179–206.

Yang, Y., Munz, J., Cass, C., Zienkiewicz, A., Kong, Q., Ma, W., et al. (2015). Ectopic expression of WRINKLED1 affects fatty acid homeostasis in brachypodium distachyon vegetative tissues. *Plant Physiol.* 169, 1836–1847.

Yuan, G. F., Chen, X. E., and Li, D. (2014). Conjugated linolenic acids and their bioactivities: a review. *Food Funct.* 5, 1360–1368.

Yuan, L., and Li, R. (2020). Metabolic engineering a model oilseed. *Front. Plant Sci.* 11, 11.

Yu, X. H., Cahoon, R. E., Horn, P. J., Shi, H., Prakash, R. R., Cai, Y., et al. (2018). Identification of bottlenecks in the accumulation of cyclic fatty acids in camelina seed oil. *Plant Biotechnol. J.* 16, 926–938.

Yu, X. H., Cai, Y., Chai, J., Schwender, J., and Shanklin, J. (2019). Expression of a lychee PHOSPHATIDYLCHOLINE : DIACYLGLYCEROL CHOLINEPHOSPHOTRANSFERASE with an escherichia coli CYCLOPROPANE SYNTHASE enhances cyclopropane fatty acid accumulation in camelina seeds. *Plant Physiol.* 180, 1351–1361.

Yu, X. H., Cai, Y., Keereetaweep, J., Wei, K., Chai, J., Deng, E., et al. (2021). Biotin attachment domain-containing proteins mediate hydroxy fatty acid-dependent inhibition of acetyl CoA carboxylase. *Plant Physiol.* 185, 892–901.

Yu, X. H., Prakash, R. R., Sweet, M., and Shanklin, J. (2014). Coexpressing escherichia coli cyclopropane synthase with sterculia foetida lysophosphatidic acid acyltransferase enhances cyclopropane fatty acid accumulation. *Plant Physiol.* 164, 455–465.

Yu, X. H., Rawat, R., and Shanklin, J. (2011). Characterization and analysis of the cotton cyclopropane fatty acid synthase family and their contribution to cyclopropane fatty acid synthesis. *BMC Plant Biol.* 11, 97.

Yurchenko, O., Shockey, J. M., Gidda, S. K., Silver, M. I., Chapman, K. D., Mullen, R. T., et al. (2017). Engineering the production of conjugated fatty acids in arabidopsis thaliana leaves. *Plant Biotechnol. J.* 15, 1010–1023.

Zale, J., Jung, J. H., Kim, J. Y., Pathak, B., Karan, R., Liu, H., et al. (2016). Metabolic engineering of sugarcane to accumulate energy-dense triacylglycerols in vegetative biomass. *Plant Biotechnol. J.* 14, 661–669.

Zhou, X. R., Singh, S. P., and Green, A. G. (2013). Characterisation of the FAD2 gene family from hiptage benghalensis: a ricinoleic acid accumulating plant. *Phytochemistry* 92, 42–48.

Zhou, X. R., Singh, S., Liu, Q., and Green, A. (2006). Combined transgenic expression of Δ12-desaturase and Δ12-epoxygenase in high linoleic acid seeds leads to increased accumulation of vernolic acid. *Funct. Plant Biol.* 33, 585–592.

Zienkiewicz, K., Zienkiewicz, A., Poliner, E., Du, Z. Y., Vollheyde, K., Herrfurth, C., et al. (2017). Nannochloropsis a rich source of diacylglycerol acyltransferases for engineering of triacylglycerol content in different hosts. *Biotechnol. Biofuels* 10, 8.

Zolman, B. K., Silva, I. D., and Bartel, B. (2001). The arabidopsis pxa1 mutant is defective in an ATP-binding cassette transporter-like protein required for peroxisomal fatty acid beta-oxidation. *Plant Physiol.* 127, 1266–1278.

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/fpls.2022.1049177

OPEN ACCESS

EDITED BY
Zhi-Yan (Rock) Du,
University of Hawaii at Manoa,
United States

REVIEWED BY
Kyle J. Lauersen,
King Abdullah University of Science and Technology, Saudi Arabia
Jörg Toepel,
Helmholtz Association of German Research Centres (HZ), Germany
Tom Bibby,
University of Southampton,
United Kingdom

*CORRESPONDENCE
Silas B. Mellor
silasmellor@plen.ku.dk

†PRESENT ADDRESS
Mathias Pribil, Enzymes &Plasma Proteins (DOZBE), Roche Diagnostics GmbH, Penzberg, Germany

SPECIALTY SECTION
This article was submitted to Plant Systems and Synthetic Biology, a section of the journal Frontiers in Plant Science

RECEIVED 20 September 2022
ACCEPTED 14 December 2022
PUBLISHED 09 January 2023

CITATION
Mellor SB, Behrendorff JBYH, Ipsen JØ, Crocoll C, Laursen T, Gillam EMJ and Pribil M (2023) Exploiting photosynthesis-driven P450 activity to produce indican in tobacco chloroplasts.
Front. Plant Sci. 13:1049177.
doi: 10.3389/fpls.2022.1049177

Exploiting photosynthesis-driven P450 activity to produce indican in tobacco chloroplasts

Silas B. Mellor[1]*, James B. Y. H. Behrendorff[2,3], Johan Ø. Ipsen[4], Christoph Crocoll[5], Tomas Laursen[1], Elizabeth M. J. Gillam[6] and Mathias Pribil[7†]

[1]Section for Plant Biochemistry, Department of Plant and Environmental Science, University of Copenhagen, Frederiksberg, Denmark, [2]School of Biology and Environmental Science, Queensland University of Technology, Brisbane, QLD, Australia, [3]Australian Research Council (ARC) Centre of Excellence in Synthetic Biology, Queensland University of Technology, Brisbane, QLD, Australia, [4]Section for Forest, Nature and Biomass, Department of Geosciences and Natural Resource Management, University of Copenhagen, Frederiksberg, Denmark, [5]DynaMo Center, Section for Molecular Plant Biology, Department of Plant and Environmental Science, University of Copenhagen, Frederiksberg, Denmark, [6]School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, QLD, Australia, [7]Section for Molecular Plant Biology, Department of Plant and Environmental Science, University of Copenhagen, Frederiksberg, Denmark

Photosynthetic organelles offer attractive features for engineering small molecule bioproduction by their ability to convert solar energy into chemical energy required for metabolism. The possibility to couple biochemical production directly to photosynthetic assimilation as a source of energy and substrates has intrigued metabolic engineers. Specifically, the chemical diversity found in plants often relies on cytochrome P450-mediated hydroxylations that depend on reductant supply for catalysis and which often lead to metabolic bottlenecks for heterologous production of complex molecules. By directing P450 enzymes to plant chloroplasts one can elegantly deal with such redox prerequisites. In this study, we explore the capacity of the plant photosynthetic machinery to drive P450-dependent formation of the indigo precursor indoxyl-β-D-glucoside (indican) by targeting an engineered indican biosynthetic pathway to tobacco (*Nicotiana benthamiana*) chloroplasts. We show that both native and engineered variants belonging to the human CYP2 family are catalytically active in chloroplasts when driven by photosynthetic reducing power and optimize construct designs to improve productivity. However, while increasing supply of tryptophan leads to an increase in indole accumulation, it does not improve indican productivity, suggesting that P450 activity limits overall productivity. Co-expression of different redox partners also does not improve productivity, indicating that supply of reducing power is not a bottleneck. Finally, *in vitro* kinetic measurements showed that the different redox partners were efficiently reduced by photosystem I but plant ferredoxin provided the highest light-

dependent P450 activity. This study demonstrates the inherent ability of photosynthesis to support P450-dependent metabolic pathways. Plants and photosynthetic microbes are therefore uniquely suited for engineering P450-dependent metabolic pathways regardless of enzyme origin. Our findings have implications for metabolic engineering in photosynthetic hosts for production of high-value chemicals or drug metabolites for pharmacological studies.

KEYWORDS

cytochrome P450, chloroplast, indigo, transient expression, photosynthesis, ferredoxin, metabolic engineering

Introduction

Biotechnological production of valuable and useful molecules is a burgeoning field, and in some cases has become the preferred method to produce biopharmaceuticals or small molecule compounds (Madhavan et al., 2021; Wu et al., 2021). Bio-based small molecule production has many environmental advantages by virtue of being carried out in aqueous systems using simple sugar or organic acid precursor molecules. This renders both feedstock and waste streams less environmentally harmful and, where extraction from natural sources is otherwise required, avoids putting undue stress on scarce or endangered natural resources. A growing number of production processes relying on metabolically engineered organism is emerging as consumer demand and interest for products from sustainable sources increases (Nielsen et al., 2022). Microbial hosts still dominate the field of engineered biomolecule production, both for protein- and small molecule-based therapeutics, but plants can be economically competitive vis-à-vis microorganisms for heterologous production (Holtz et al., 2015). Plants also offer the enticing possibility of coupling biochemical production directly to the assimilation of light energy and carbon *via* photosynthesis; truly green chemistry (Sørensen et al., 2022). Recent thorough techno-economic analyses have shown that bio-fuel manufacturing in photosynthetic algae can be economically viable given optimal environmental conditions, especially if co-products such as protein can be extracted in parallel from the same strain (Roles et al., 2021; Karan et al., 2022). This offers hope that with proper design and optimization, production of high value molecules through metabolic engineering in plants could prove economically feasible in the future.

Most examples of plants engineered to produce small molecules have introduced pathways in the cytosol, but compartmentation of pathways is a growing trend (Huttanus and Feng, 2017), and follows an increased understanding of the compartmentalized nature of many biosynthetic pathways (Heinig et al., 2013). Chloroplasts serve as hubs for cellular bioenergy generation by virtue of their primary photosynthetic processes and their role as a carbon fixation, storage and re-mobilization compartment. Chloroplasts are also highly diverse metabolic centers that supply precursors for a wide variety of central and specialized metabolic processes (Heinig et al., 2013; Nielsen et al., 2016). As such they are attractive compartments for introducing a variety of metabolic pathways, such as those dependent on oxyfunctionalization by cytochrome P450 enzymes (Mellor et al., 2016). Because of the widespread occurrence of P450s throughout all kingdoms of life, their prominent roles in specialized metabolism and ability to catalyze remarkably diverse chemical transformations (Podust and Sherman, 2012; Liu et al., 2020; Behrendorff, 2021), they are important enzymes in the metabolic engineering toolbox. Most P450s in eukaryotes localize to the endoplasmic reticulum (ER) where they are anchored by a single N-terminal transmembrane domain. Eukaryotes likewise harbor distinct diflavin cytochrome P450 reductases (CPRs), which are also membrane bound and localize to the ER confining metabolism to the two-dimensional membrane lattice (Laursen et al., 2021). Being membrane proteins and dependent on dedicated reductase systems, P450s are often difficult to functionally reconstitute in heterologous hosts and may require significant engineering effort to reach acceptable functionality (Renault et al., 2014; Liu et al., 2020; Jensen et al., 2021).

We and others have shown heterologous targeting of P450s to tobacco chloroplasts by fusion to the chloroplast targeting peptide from *Arabidopsis* ferredoxin 2 and found that photosynthesis supplies reducing equivalents to the P450s *via* interaction with chloroplast ferredoxins (Nielsen et al., 2013; Mellor et al., 2016). Targeting P450 enzymes to thylakoid membranes and avoiding the need for co-expression of a dedicated reductase is an advantage for engineering of P450-dependent pathways but has only been shown for a handful of P450s to date (Nielsen et al., 2013; Xue et al., 2014; Gnanasekaran et al., 2015; Berepiki et al., 2016; Fräbel et al., 2018; Li et al., 2019). Several human P450s produce the insoluble blue dye indigo upon recombinant expression in *E. coli* (Gillam et al., 1999). Expression of the

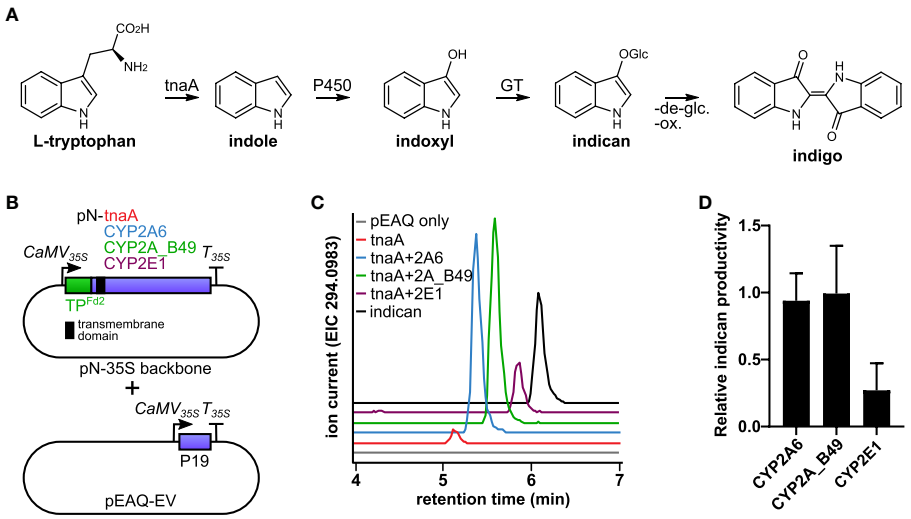


FIGURE 1
(A) Diagram of the indigo pathway introduced into tobacco chloroplasts in this study. Tryptophan is converted by tnaA to indole, which is hydroxylated to indoxyl by P450s and subsequently glycosylated by endogenous tobacco glucosyltransferases (GT) to form indican. Indigo can be formed from indican by de-glycosylation and oxidative polymerization. **(B)** Schematic of constructs used to establish indican production in chloroplasts using human CYP2 isoforms. Both tnaA and P450s were targeted to chloroplast by *Arabidopsis* Fd2 transit peptide (TP^{Fd2}). The P450s inserted into thylakoid membranes by their N-terminal transmembrane (indicated by a black square). *Agrobacterium* carrying vector encoding tnaA were infiltrated on its own or in combination with vector carrying a human CYP2 isoform, as indicated at equal ODs together with empty pEAQ vector (pEAQ-EV) to deliver the P19 suppressor of silencing protein. **(C)** Production of indican by CYP2A6, CYP2A_B49, and CYP2E1. Extracted ion chromatograms from LC-QTOF-MS/MS analysis of the $[M-H]^-$ molecular ion of indican (*m/z* 294.0983). **(D)** Relative productivity of CYP2A6, CYP2A_B49 and CYP2E1 co-expressed transiently in tobacco together with *E. coli* tnaA, determined by comparison of internal standard-normalized indican peak areas from 10-11 individually infiltrated tobacco leaves (error bars show SD).

indole hydroxylase CYP2A6 in tobacco together with either maize indole-3-glycerol phosphate lyase (BX1) or *E. coli* tryptophanase (tnaA) results in accumulation of the glucosylated indigo precursor indoxyl β-D-glucoside (Figure 1A, indican) (Warzecha et al., 2007; Fräbel et al., 2018). Here we test the ability of tobacco chloroplasts to support human P450s by targeting three indole hydroxylating P450 isoforms and tnaA to tobacco chloroplasts and use the system to explore strategies for optimizing photosynthesis-driven indican productivity. We find that maximal productivity is achieved by expressing P450 and tnaA from the same vector and co-express a feedback insensitive mutant of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (AroG*), which catalyzes the first committed step of the shikimate pathway to boost the supply of tryptophan. Its transient expression in tobacco yields 37 to 115-fold increases in aromatic amino acid pools, with concurrent effects on indole production by tryptophanase. We also co-express different redox partners but find that neither these nor AroG* affect indican productivity. Through *in vitro* study on isolated electron transfer proteins, we find that plant ferredoxin outcompetes other redox partners in coupling photosynthetic electron transport and P450 activity. Together, our results show that P450 activity constitutes the major bottleneck for overall indican productivity in our transient expression system.

Materials and methods

Expression vectors for *E. coli* expression

The pTrc99a-AtFd1 vector was a kind gift from Guy T. Hanke. Flavodoxin genes were amplified by PCR with overhangs and cloned into pET14b vector using NEB HiFi DNA assembly master mix (New England Biolabs). The IsiB gene from *Synechocystis* sp. PCC6803 was amplified by colony PCR. The flavodoxin-like domain of human CPR (residues 62-241, referred to as $HsCPR^{62-241}$) was amplified from the bicistronic expression vector pCWori/CYP2A_B49/hNPR (Strohmaier et al., 2019). After assembly, colony PCR positive clones were confirmed by sequencing.

Expression vectors for transient expression in *N. benthamiana*

The CYP2A_B49 variant was selected based on indigo productivity in *E. coli* generated from a DNA shuffling library using CYP2A5, CYP2A6, and CYP2A13 isoforms (Behrendorff et al., 2013; Strohmaier et al., 2019). Cytochrome P450 genes were amplified from *E. coli* expression constructs (Gillam et al.,

1999). TnaA and AroG* genes were synthesized by Twist Bioscience (San Francisco, USA). All inserts were PCR amplified with overhangs and were cloned into the pN vector (Karimi et al., 2002; Behrendorff et al., 2019) using NEB HiFi DNA assembly master mix (New England Biolabs). For bidirectional promoter constructs, the pN vector was used without 35S promoter and insert cassette. AtUBQ10 promoter was amplified from a previously reported construct (Behrendorff et al., 2019). *Nicotiana tabacum* ferredoxin chloroplast transit peptide and *Arabidopsis thaliana* Hsp18.2 terminator were synthesized by Twist Bioscience (San Francisco, USA). The Ubi.U4 promoter was amplified from *N. tabacum* genomic DNA. After assembly, colony PCR positive clones were confirmed by sequencing. pEAQ-HT vectors with AroG*, AtFd1, $IsiB^{6803}$, $HsCPR^{62-241}$ and $HsCPR^{fl}$ inserts were synthesized by Twist Bioscience (San Francisco, USA). Nucleotide sequences of the genes, promoters and terminators used are given in the supplementary materials (Supplementary Table S1).

Transient expression in *N. benthamiana*

Colonies of *Agrobacterium tumefaciens* GV3101 transformed with vectors of interest were inoculated in YEP in the presence of 10 μg mL^{-1} rifampicin, 25 μg mL^{-1} gentamycin with 100 μg mL^{-1} spectinomycin (pN, pBJ1) or 50 μg mL^{-1} kanamycin (pEAQ) and grown with shaking (200 rpm) for 2 d at 28°C. Cells were harvested and suspended in infiltration buffer (10 mM MES pH 5.6, 10 mM $MgCl_2$, 200 μM acetosyringone) to the desired OD and incubated at RT with gentle agitation for 2 h. Leaves of 4-6 week old greenhouse-grown tobacco plants were infiltrated on the abaxial leaf surface. Plants were allowed to dry before returning to greenhouse.

Untargeted LC-QTOF-MS/MS

Samples (80-120 mg) were taken from *N. benthamiana* plants 3-5 days post inoculation (dpi), snap frozen in liquid N_2 and homogenized using a Retsch Tissue-Lyzer mixer mill (2-4x30 s, 20 s^{-1}). The homogenized material was extracted with 0.5 mL 80% methanol for 30 min at 4°C. Samples were centrifuged (2,250 g, 10 min, 4°C) and supernatant diluted 4-fold into milliQ water containing 1 mg L^{-1} Leu-enkephalin as internal standard. LC-MS/MS was performed on a Dionex UltiMate 3000 Quaternary Rapid Separation UHPLC+ focused system (Thermo Fisher Scientific, Germering, Germany) using a Kinetex XB-C18 column (100 × 2.1 mm, 1.7 μm, 100 Å, Phenomenex). Formic acid (0.05% v/v) in water and acetonitrile (supplied with 0.05% v/v formic acid) were employed as mobile phases A and B, respectively. Gradient conditions were as follows: 0.0-0.5 min 5% B; 0.5–.0 17.5 min 5–45% B, 17.5-23.0 min 45-75% B, 23.0-25.0 min 75-100% B, 25.0–26.95 min 100% B, 26.95-27.0 min 100–5% B, and 27.0 –30.0 min 5% B. The flow rate of the mobile phase was 300 μL min^{-1}. The column temperature was maintained at 30°C. UV spectra for each sample were acquired at 214, 260, 280, and 300 nm. The UHPLC was coupled to a Compact micrOTOF-Q mass spectrometer (Bruker, Bremen, Germany) equipped with an electrospray ion source (ESI) operated in positive or negative ion mode. The ion spray voltage was maintained at +4500 V and -3900 V in positive and negative ion mode, respectively. Dry temperature was set to 250°C, and nitrogen was used as the dry gas (8 L min^{-1}), nebulizing gas (2.5 bar), and collision gas in both ion modes. Collision energy was set to 10 eV and 15 eV in positive and negative ion mode, respectively. MS spectra were acquired with a sampling rate of 2 Hz in an *m/z* range from 50 to 1000 and MS/MS spectra in a range of *m/z* 100-800 in positive ion mode while in negative ion mode MS spectra were acquired with a sampling rate of 3 Hz in an *m/z* range from 50 to 1400 and MS/MS spectra in a *m/z* range from 100-1000. Na-formate clusters were used for mass calibration in both ion modes. Positive ion mode was used for untargeted metabolomics and negative ion mode was used to detect indican. Indole did not provide detectable MS signal and was instead quantified using its UV absorbance at 214 nm against a standard curve containing 0.75-100 μM indole in 20% methanol.

Analysis of untargeted MS data

Data analysis was performed using MS-DIAL (Lai et al., 2017). A MS1 tolerance of 0.05 Da was used, and features were matched to the positive ionization LC-MS/MS library from Mass Bank of North America (https://mona.fiehnlab.ucdavis.edu/) using MS1 and MS2 tolerances of 0.01 Da and 0.025 Da, respectively. After alignment, data was normalized to the peak area of the Leu-Enk internal standard. PCA analysis was performed on log_{10}-transformed and auto-scaled data using only reference matched features. Normalized peak areas of compounds of interest were extracted directly from MS-DIAL and plotted using GraphPad Prism.

Quantification of indican by LC-QqQ-MS/MS

Samples were collected at 5 dpi, homogenized and extracted as for untargeted MS analysis. Extracts were diluted 20 times into 20% methanol with 50 μg L^{-1} *p*-hydroxybenzaldehyde as internal standard and subjected to LC-MS analysis. Chromatography was performed on a 1290 Infinity II UHPLC system (Agilent Technologies), using a Kinetex XB-C18 column (100 x 2.1 mm, 1.7 μm, 100 Å, Phenomenex, Torrance, CA, USA). Formic acid (0.05%, v/v) in water and acetonitrile

(supplied with 0.05% v/v formic acid) were employed as mobile phases A and B, respectively. The elution profile for indican was: 0.0-2.5 min, 5-65% B; 2.5-3.0 min 65-100% B, 3.0-4.0 min 100% B, 4.0-4.1 min, 100-5% B and 4.1-5.0 min 5% B. The mobile phase flow rate was 400 µL min^{-1}. The column temperature was maintained at 40°C. The liquid chromatography was coupled to an Ultivo Triplequadrupole mass spectrometer (Agilent Technologies) equipped with a Jetstream electrospray ion source (ESI) operated in negative ion mode. The ion spray voltage was set to -3000 V. Dry gas temperature was set to 325°C and dry gas flow to 9 L min^{-1}. Sheath gas temperature was set to 325°C and sheath gas flow to 12 L min^{-1}. Nebulizing gas was set to 40 psi. Nitrogen was used as dry gas, nebulizing gas and collision gas. Multiple reaction monitoring (MRM) was used to monitor precursor ion → fragment ion transitions (Supplementary Table S2). MRM transitions and instrument parameters were optimized using reference standards. Both Q1 and Q3 quadrupoles were maintained at unit resolution. Mass Hunter Quantitation Analysis for QqQ software (Version 10.1, Agilent Technologies) was used for data processing. Indican peak areas were converted to concentrations by comparing against a standard curve ranging from 0.025-5 µM. Standard curves of authentic indican (abcam) were prepared in blank tobacco matrix prepared as described above from untransformed *N. benthamiana* to account for matrix effects.

Quantification of amino acids by LC-QqQ-MS/MS

Samples were collected 5 dpi as described above but extracted using 85% methanol. Subsequently, samples were mixed 1:10 (v/v) with ^{13}C, ^{15}N labeled amino acids (Algal amino acids ^{13}C, ^{15}N, Isotec, Miamisburg, US) at a concentration of 10 µg*mL^{-1} either directly or after 10-fold dilution with water and analyzed by LC-MS. The analysis was performed as described in (Mirza et al., 2016) with changes as detailed below. Chromatography was performed on an Advance UHPLC system (Bruker, Bremen, Germany) using a Zorbax Eclipse XDB-C18 column (100 x 3.0 mm, 1.8 µm, Agilent Technologies, Germany). Formic acid (0.05% v/v) in water and acetonitrile (supplied with 0.05% v/v formic acid) were employed as mobile phases A and B, respectively. The elution profile was: 0-1.2 min 3% B; 1.2-4.3 min 3-65% B; 4.3-4.4 min 65-100% B; 4.4-4.9 min 100% B, 4.9-5.0 min 100-3% B and 5.0-6.0 min 3% B. Mobile phase flow rate was 500 µL min^{-1} and column temperature was maintained at 40°C. The liquid chromatography was coupled to an EVOQ Elite TripleQuad mass spectrometer (Bruker, Bremen, Germany) equipped with an electrospray ionization source (ESI). Instrument parameters were optimized by infusion experiments with pure standards. The ion spray voltage was maintained at 3000 V in positive ion mode. Cone temperature was set to 300°C and cone gas flow to 20 psi. Heated probe temperature was set to 400°C and probe gas flow set to 50 psi. Nebulizing gas was set to 60 psi and collision gas to 1.6 mTorr. Nitrogen was used as both cone gas and nebulizing gas and argon as collision gas. Analyte precursor ion → fragment ion transitions were monitored by MRM, with transitions chosen as described by Jander et al., with additions from Docimo et al. for Arg and Lys (Jander et al., 2004; Docimo et al., 2012). Both Q1 and Q3 quadrupoles were maintained at unit resolution. Bruker MS Workstation software (Version 8.2.1, Bruker, Bremen, Germany) was used for data acquisition and processing. Individual amino acids in the sample were quantified by comparison of peak areas of light- amino acids with ^{13}C, ^{15}N-labeled amino acid internal standards of known concentrations, except for tryptophan, asparagine and glutamine. Tryptophan was quantified using ^{13}C, ^{15}N-Phe applying a response factor of 0.42, asparagine and glutamine were quantified using ^{13}C, ^{15}N-Asp and ^{13}C, ^{15}N-Glu, respectively, applying a response factor of 1.0 (Docimo et al., 2012).

Overexpression and purification of electron transfer proteins

AtFd1 was expressed in *E. coli* DH5α after inducing cells with 1 mM Isopropyl *β-D*-1-thiogalactopyranoside (IPTG) at OD=1.0 and growing the cells overnight at 37°C with 200 rpm shaking. Purification was performed essentially as previously described (Matsumura et al., 1999). Flavodoxins were expressed in *E. coli* BL21 (DE3) after inducing cells with 1 mM IPTG at OD=0.5-0.6 and growing the cells overnight at 30°C ($IsiB^{6803}$) or 25°C ($HsCPR^{62-241}$). Cells were and harvested by centrifugation (10,000 g for 30 min at 4°C) and stored at -20°C. Cells were lysed by incubation for 30 min on ice with stirring in lysis buffer (50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 mM imidazole, 0.5 mg mL^{-1} lysozyme) before sonicating for 15 cycles (30 s, 50% duty cycle, amplitude 6 followed by 1 min cooling period) using a Branson Sonifier 450 sonicator. Lysates were clarified by centrifugation (14,000 g, 20 min, 4°C) followed by filtering at 0.45 µm. Purification was performed using a Cytiva HisTrap HP column at 1 mL min^{-1} on an ÄKTA Start purification system (Cytiva). The column was washed (20 column volumes, 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 5 mM imidazole) prior to eluting with a linear gradient of 5-250 mM imidazole over 30 column volumes. Yellow fractions were pooled, desalted into 50 mM Tris-HCl (pH 7.5) and checked for purity using SDS-PAGE.

Reconstitution of flavodoxins and determination of FMN extinction coefficients

Flavodoxins (~1-2 mM) were reconstituted with 10 mM FMN on ice overnight and desalted into 50 mM tris-HCl (pH

7.5). FMN incorporation was estimated by comparing with calculated absorbance ratios for FMN and protein at 280 nm (0.166, and 0.210, respectively for $IsiB^{6803}$ and $HsCPR^{62-241}$). FMN incorporation was typically >95%. Flavodoxin extinction coefficients were determined by adapting established protocols (Mayhew and Massey, 1969) for microtiter plates. FMN cofactor was extracted by successive treatments with 5% TCA on ice, the supernatants recovered after centrifugation at 10,000 g for 10 min and neutralized by adding K_2HPO_4 to 0.3 M. FMN extinction coefficients were calculated as $\epsilon = 12.2\ mM^{-1}\ cm^{-1} * (A_{FMN}{}^{bound}/A_{FMN}{}^{free})$, based on spectrophotometric measurements of FMN extracted from flavodoxins and holoflavodoxins measured in UV-clear microtiter plates using a BioTek Synergy H1 plate reader and corrected for buffer background. Coefficients of 8.7 and 10.1 $mM^{-1}\ cm^{-1}$ were obtained, respectively, for $IsiB^{6803}$ and $HsCPR^{62-241}$.

Isothermal titration calorimetry

Isothermal titration calorimetry was performed on a MicroCal PEAQ-ITC calorimeter (Malvern Panalytical). Titrations were performed by injecting FMN (>95% purity, Merck Sigma-Aldritch) into a solution of pure $IsiB^{6803}$ or $HsCPR^{62-241}$ apoprotein desalted into 50 mM Tris-HCl pH 7.5. The same preparation of buffer was used for both protein and ligand. Protein concentration in the cell was 15 µM and titrant used was 150 µM. After a priming injection of 0.4 µL titration was performed using a total of 12 injections of 3 µL and the resulting data was fitted using the MicroCal PEAQ-ITC analysis software (Malvern Panalytical).

Thylakoid preparation

Thylakoids were prepared by homogenizing leaves of dark-adapted tobacco plants thoroughly in ice-cold homogenization buffer (0.4 M sucrose, 20 mM Tricine-NaOH pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, 100 mM sodium ascorbate, 5 mg mL^{-1} bovine serum albumin) using a blender. The homogenate was filtered centrifuged for 10 min at 5,000 g, 4°C. Pelleted chloroplasts were resuspended in 5 mM Tricine pH 7.9 and left to rupture osmotically on ice for 15 min. Thylakoid membranes were pelleted at 11,200 g, 4°C for 10 min and resuspended in storage buffer (0.4 M sucrose, 20 mM Tricine-NaOH pH 7.5, 10 mM NaCl, 5 mM $MgCl_2$, 20% v/v glycerol). Chlorophyll concentration of preparations was measured according to Lichtenthaler (Lichtenthaler, 1987).

Light-dependent coumarin 7-hydroxylation assays

The activity of CYP2A_B49 was measured after isolating thylakoid membranes from pN-CYP2A_B49-infiltrated tobacco plants. Assays contained 50 mM tricine (pH 7.5), 100 mM NaCl, 10 µM coumarin, CYP2A_B49-containing thylakoids (100 µg chl mL^{-1}) and redox partners (0.1-30 µM). Reactions were maintained at 25°C and were started by illuminating assays in black 96-well microtiter plate using a rectangular LED array connected to a laboratory power supply and adjusted to deliver a photon flux of 170 µmol $m^{-2}\ s^{-1}$ light intensity. Reactions were stopped after 2-10 min by addition of 0.33 µM methyl viologen and switching off the light. Plates were centrifuged (2,850 g, 5 min) and supernatants were filtered through 0.45 µm filters to remove thylakoid membranes. Umbelliferone fluorescence was measured against authentic standards (5-100 nM) dissolved in 50 mM tricine (pH 7.5), using excitation and emission wavelengths of 370 and 450 nm, respectively. Enzymatic rates were corrected for background rates without redox partner and the data fitted to the Michaelis-Menten equation to obtain kinetic parameters using GraphPad Prism.

Cytochrome c reduction assays

Light-dependent cytochrome *c* reduction was measured using a Shimadzu UV-2550 double beam spectrophotometer fitted with a light fiber and 550 nm band pass filters on the measuring beam windows. Temperature was controlled at 25°C by thermostat. Assays contained 50 mM Tricine (pH 7.5), 100 mM NaCl, 200 µM cytochrome *c*, thylakoids (10 µg chl mL^{-1}) and 0.01-4 µM redox partners. The assay was initiated by switching on the light source (Schott KL-1500 fitted with red-light bandpass filters) and the activity was followed for 60 s. Initial rates of cytochrome *c* reduction were calculated by taking a reduced-oxidized extinction coefficient of 18.5 $mM^{-1}\ cm^{-1}$ (Kubota et al., 1992), corrected for the rate of cytochrome *c* reduction without redox partner present and fitted to the Michaelis-Menten equation using GraphPad Prism.

Ferredoxin : NADP+ reductase assays

Assays were carried out at RT in 96-well microtiter plates by adding 50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 1 mM $MgCl_2$, 5 mM glucose 6-phosphate, 5 U mL^{-1} glucose 6-phosphate dehydrogenase, 200 µM cytochrome *c*, and 40 nM spinach FNR to microtiter wells containing NADPH (500 µM final) and electron carriers (1-100 µM final). Absorbance was

measured at 550 nm for 2 min using a BioTek Syntergy H1 plate reader. Initial rates of cytochrome *c* reduction were calculated taking a reduced-oxidized extinction coefficient of 18.5 mM^{-1} cm^{-1} (Kubota et al., 1992) and corrected for the background cytochrome *c* reduction rate in assays without electron carriers. The data was fitted to the Michaelis-Menten equation using GraphPad Prism.

Results

Photosynthesis-driven indican biosynthesis using human cytochrome P450 enzymes

Chloroplasts were previously shown to provide membrane insertion and light-dependent supply of reducing power towards catalysis by plant cytochrome P450 enzymes (Jensen et al., 2011; Nielsen et al., 2013). We wanted to investigate whether this principle could be extended to support catalytic activity by distantly related P450 enzymes such as those from animals. We therefore expressed human P450s CYP2A6 and CYP2E1 and the engineered variant CYP2A_B49, all of which hydroxylate indole to indoxyl (Gillam et al., 1999) transiently targeted to chloroplasts in *Nicotiana benthamiana*, and used the hydroxylation of indole as a readout for P450 activity (Figure 1A). Indole hydroxylation was detectable as accumulation of indoxyl β-D-glucoside, the major hydroxylated indole metabolite that occurs in plants as a result of indoxyl glycosylation by endogenous plant glucosyltransferases (Warzecha et al., 2007). Since tryptophan biosynthesis occurs in chloroplasts, we co-expressed *E. coli* L-tryptophan indole lyase (tnaA), which hydrolyses tryptophan to indole, also with chloroplast targeting to provide indole precursor for the P450s. We used the chloroplast transit peptide from *Arabidopsis* ferredoxin 2 (TP^{Fd2}), which can direct both membrane targeted P450s and soluble enzymes to chloroplasts in tobacco (Nielsen et al., 2013), fused at the N-terminus of both P450s and tnaA to target the enzymes to chloroplasts (Figure 1B). *Agrobacterium* strain transfected with empty pEAQ-HT vector (Peyret and Lomonossoff, 2013) was mixed with pN-carrying strains at equal ODs in all transfections to reduce silencing and ensure consistent maximal expression of tnaA and P450s. Expression of tnaA alone led to detectable indole accumulation in leaf samples (Supplementary Figure S1). We detected trace amounts of indican when tnaA was expressed alone, possibly due to presence of endogenous indole-metabolizing enzymes, as reported previously (Warzecha et al., 2007). However accumulation was considerably higher when any P450 was co-expressed with tnaA (Figure 1C). Four additional unknown compounds were detected upon expression of tnaA with CYP2A6 or CYP2A_B49 (Supplementary Figure S2A). Three of them had identical parent masses and were consistent with dioxyindole glucosides (Supplementary Figure S2B), while the parent mass of the fourth compound was consistent with hydroxylated indole acetic acid glucoside (Supplementary Figure S2C). Co-infiltration of tnaA with CYP2A6 and CYP2A_B49 yielded similar indican accumulation (Figure 1D), and CYP2A_B49 was used for subsequent experiments.

Optimization of transient indican production by T-DNA engineering

Indican production requires both tnaA and P450 transgenes and we hypothesized that placing both transgenes on the same T-DNA would improve indican productivity by ensuring all transfected cells receive both transgenes. To test this hypothesis, we cloned two bi-directional pN constructs (pBJ1-1 and pBJ1-2) that contained both tnaA and CYP2A_B49 genes driven by different promoters (Figure 2A). The CYP2A_B49 gene was driven by the same 35S promoter/terminator combination as our single gene pN vectors in both pBJ1 versions, whereas tnaA expression was driven either by the *Arabidopsis* polyubiquitin 10 promoter (At_{UBQ10}, pBJ1-1) or *N. tabacum* polyubiquitin U4 promoter (Kang et al., 2003) ($Nt_{UBQ.U4}$, pBJ1-2), both with the *Arabidopsis* Hsp18.2 terminator (Nagaya et al., 2010). We also used *N. tabacum* ferredoxin CTP to avoid re-using that of *Arabidopsis* Fd2 to target tnaA to chloroplasts. Having both enzymes on the same T-DNA nearly doubled indican accumulation compared to mixing individual *Agrobacterium* strains, with the pBJ1-1 and pBJ1-2 constructs producing equivalent indican amounts (Figure 2B). These results indicate that expressing multiple interdependent transgenes from the same T-DNA can increase productivity compared to introducing the transgenes on individual vectors.

Engineering the shikimate pathway to boost aromatic amino acid accumulation

We next decided to investigate if indican production could be further improved by boosting supply of substrates to CYP2A_B49. Because the indole substrate for CYP2A_B49 derives from conversion of tryptophan by tnaA, we hypothesized that increasing aromatic amino acid supply might increase indican production in our system. We had previously observed increased MS peak areas belonging to Phe, Tyr and Trp (not shown) after transient expression of a feedback insensitive variant of 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase, $AroG^{D146N/A202T}$ (Ding et al., 2014) - denoted AroG* in the following. To quantify this increase, we expressed AroG* targeted to chloroplasts with the CTP from AtFd2 using our pN vector together with pEAQ-EV and performed absolute quantification of 18 amino acids (Ala, Arg,

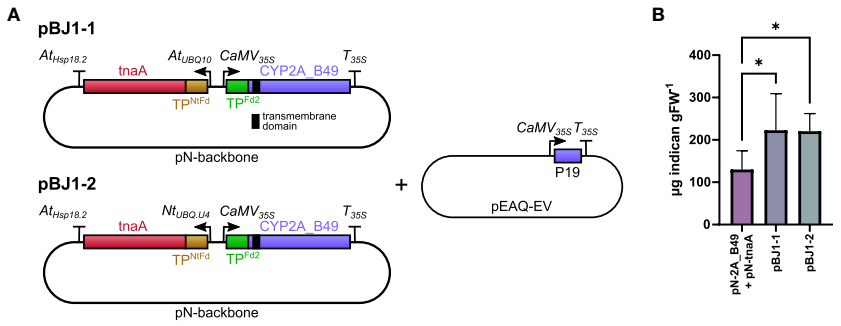


FIGURE 2
(A) Schematic of pBJ bidirectional promoter T-DNA constructs used in this experiment. The transmembrane domain of CYP2A_B49 is indicated by a black bar. Co-expression tnaA and CYP2A_B49 was done either by co-infiltrating pN-tnaA and pN-CYP2A_B49 vectors or by infiltrating the bi-directional T-DNA pBJ1-1 or pBJ1-2 vectors at equal ODs together with pEAQ-EV to supply the P19 suppressor of silencing gene. **(B)** Indican accumulation resulting from expressing tnaA and CYP2A_B49 from individual pN vectors or together from pBJ1-1 or pBJ1-2 vectors. Bars show averages from 6 individually infiltrated tobacco leaves of similar age with error bars showing standard deviation. Asterisks indicate statistically significant differences ($p < 0.05$) in *post-hoc* pairwise Holm-Šídák tests following one-way ANOVA.

Asn, Asp, Glu, Gln, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val) at 5 dpi. Plants expressing AroG* accumulated significantly more aromatic amino acids, with 115-, 49- and 37-fold increased accumulation of Phe, Tyr and Trp, respectively (Figure 3, Table 1). Expression of AroG* also caused significantly increased accumulation of Met, Val, Ile, Leu, His and Arg, but decreased Asp and Glu accumulation, albeit to a lesser extent (Figure 3, Supplementary Table S3). Total amino acid accumulation roughly doubled upon expression of AroG* compared to empty vector controls (Supplementary Table S3).

Increased Trp or redox partner supply does not improve transient indican production

We wanted to test whether increasing Trp content would yield a corresponding increase in indican accumulation. An added benefit of the pBJ1-1 and -2 constructs is that we can co-express additional proteins by swapping them into the empty cloning site on the pEAQ-EV vector we co-infiltrate to deliver P19 (Figure 4A). This allows us to keep the total number of T-DNAs introduced constant and ensures consistent tnaA and CYP2A_B49 expression levels in each comparison. Upon co-infiltration of pEAQ-AroG* with pBJ1-1 we unexpectedly saw a drop in indican productivity compared to pBJ1-1 with pEAQ-EV (Figure 4B). To investigate whether this drop might be caused by an overall reduction in the activity of our pathway, we compared the accumulation of indole in our control (EV) and AroG* co-infiltrations by UPLC-UV. When co-infiltrating pBJ1-1 with AroG* we saw approximately four times more indole than with pEAQ-EV (Figure 4C). We also carried out untargeted MS analysis to investigate whether any other metabolite changes could explain the reduction in indican observed. Samples from co-expression of AroG* were clearly separated from EV control in PCA analysis (Supplementary Figure S3A). The main drivers of the separation were the aromatic amino acids as well as de-aminated and de-carboxylated Trp derivatives (Supplementary Figure S3B). Four compounds spectrally matched to 2-oxindoles mirrored the accumulation pattern observed for indican (Figure 4B, Supplementary Figure S4). Two of the features had retention times that coincided with indican, and probably arose from in-source fragmentation.

Next, we tested whether increasing the supply of reducing power to CYP2A_B49 might improve indican biosynthesis. We previously showed that it is possible to insert redox partners into the photosynthetic electron transfer chain at the photosystem I-stroma interface to alter the distribution of photosynthetically generated reducing power within the chloroplast. An engineered soluble flavodoxin-like protein derived from plant CPR could thus interact with and accept electrons directly from PSI (Mellor et al., 2019). To test whether the same could hold for our indican biosynthetic system, we introduced additional redox partners. As for AroG* co-expression, redox partner expression constructs were encoded on the pEAQ vector to avoid introducing additional T-DNAs into the infiltration mixture. We chose three soluble redox partners, namely AtFd1, which we have previously shown to be the ferredoxin best able to support light-driven activity of *Sorghum* CYP79A1 (Mellor et al., 2019), the cyanobacterial flavodoxin $IsiB^{6803}$ from *Synechocystis* sp. PCC6803, and the flavodoxin-like domain from human CPR, $HsCPR^{62-241}$, truncated to produce an independent soluble flavodoxin-like protein similar to the flavodoxin-like protein we previously generated from *Sorghum* CPR2b (Estrada et al., 2016; Mellor et al., 2019). We also wanted to test whether

10.3389/fpls.2022.1049177

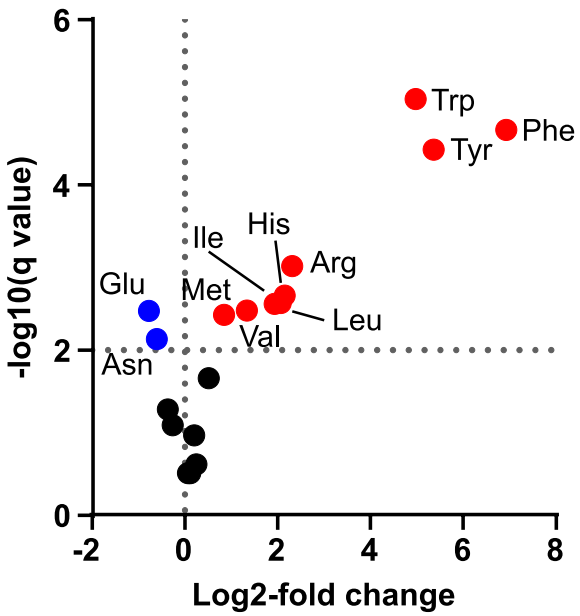


FIGURE 3
Changes in amino acid concentrations after transient expression of AroG*. Volcano plot shows the log2-fold difference in absolute amino acid abundance between plants infiltrated with pEAQ-EV alone or with pEAQ-EV and pN-AroG* together at equal OD quantified 5 dpi. Statistical significance was determined by unpaired t-tests with correction for multiple comparisons using two-stage step-up method of Benjamini, Krieger and Yekuteli (Benjamini et al., 2006), with a false discovery cutoff of 1% (*q = 0.01*, horizontal dashed line). Significantly increased amino acids are plotted in red, significantly decreased amino acids are plotted in blue and amino acids for which no significant difference was found in black.

introducing the full-length CPR, and thus moving competition for reducing power from the level of ferredoxin to NADPH (Figure 4A) would improve overall activity. When we co-infiltrated pEAQ vector carrying either redox partner we saw a similar productivity as in AroG* co-expression, except for co-expression of HsCPR^{fl}, in which case indican accumulation was about 25% that of the EV control (Figure 4B). These results show that productivity is not limited by availability of indole or redox supply, and that expressing additional transgenes from the pEAQ vector has detrimental effects on indican yield.

In vitro comparison of ferredoxin and flavodoxin redox partners

To better understand how effectively different redox partners support CYP2A_B49 catalytic activity, we next expressed and purified the soluble redox carriers AtFd1, IsiB^{6803}, and HsCPR$^{62-241}$. All purified proteins showed UV-Visible spectra characteristic of their 2Fe-2S (AtFd1, Supplementary Figure S4A) or FMN cofactors (IsiB^{6803} and HsCPR$^{62-241}$, Supplementary Figure S4B). Although both purified flavodoxins bound FMN, we performed isothermal titration calorimetry (ITC) to ensure that FMN binding was not affected by the truncation. Both flavodoxins exhibited native-like FMN binding (Supplementary Figure S4), with affinities (Supplementary Table S4) generally consistent with those reported for cyanobacterial flavodoxins and FMN-binding domains of eukaryotic CPR (Lian et al., 2011; Lans et al., 2012). Together these data show that truncated flavodoxin-like HsCPR$^{62-241}$ retains the ability to bind FMN in a native-like manner.

To examine the ability of purified redox partners to support photosynthesis-driven CYP2A_B49 activity we first measured the kinetics of electron transfer from photosystem I to each electron carrier. IsiB^{6803} outperformed the other electron carriers, with V_{max}/K_M ratio (713) almost double that of AtFd1 (420), and nearly four times that of HsCPR$^{62-241}$ (198), though all carriers showed sub-μM K_M values (Figure 5A, Table 2). We then assayed kinetics of NADPH-dependent reduction of electron carriers by FNR to investigate how strongly each electron carrier would interact with competing electron sinks present in chloroplasts. Unexpectedly, V_{max}/K_M = 13.6 for HsCPR$^{62-241}$ was about double that of AtFd1 (5.4), and 20-times higher than that of IsiB^{6803} (Figure 5B, Table 2). Finally, we measured steady state kinetics of CYP2A_B49 embedded in photosynthetic thylakoid membranes purified from tobacco after transient expression with each electron carrier protein. Because indoxyl rapidly oxidizes to form insoluble indigo dye *in vitro*, we instead adapted a commonly

TABLE 1 Average absolute quantities and ratios of aromatic amino acids in tobacco plants (*n=7* leaves from individual plants per condition) after transient expression of pEAQ-EV or pEAQ-EV + pN-AroG* constructs. The asterisk is part of the name used for the aroG* gene in this study.

Amino acid	pEAQ-EV	pEAQ-EV + pN-AroG*	Ratio (AroG*/EV)
	nmol gFW^{-1}		
Phenylalanine	12.0	1374.0	114.8
Tyrosine	5.8	286.3	49.4
Tryptophan	0.6	21.3	36.7

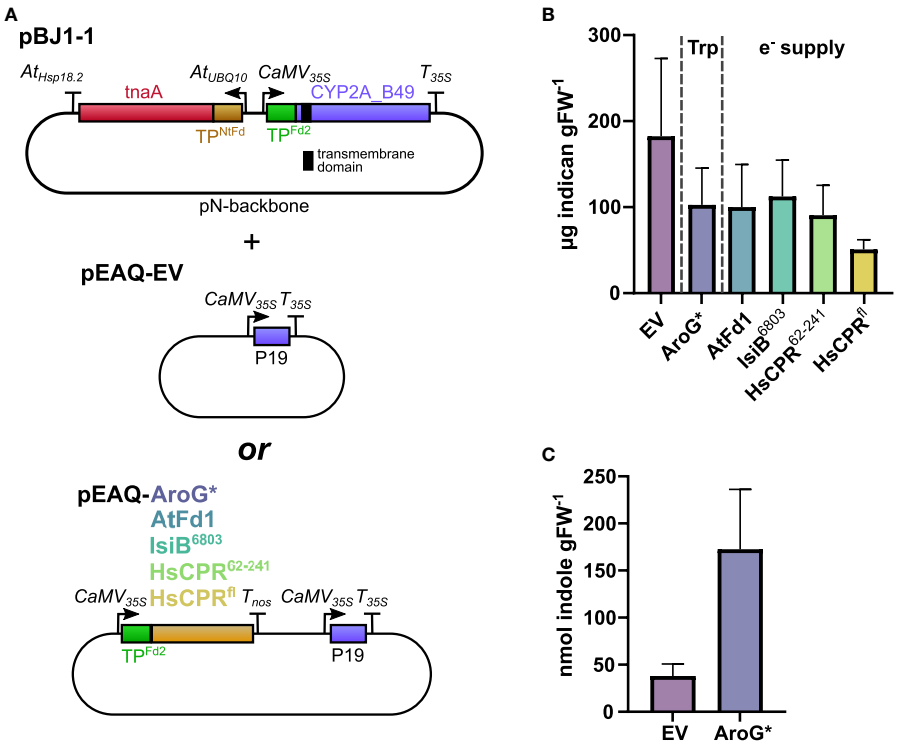


FIGURE 4
(A) Schematic of vector combinations used in this experiment. The transmembrane domain of CYP2A_B49 is indicated by a black bar. pBJ1-1 vector was co-infiltrated with either pEAQ-EV or a pEAQ containing AroG* or one of four redox partners. **(B)** Indican productivity upon co-infiltrating pBJ1-1 with pEAQ constructs carrying different transgenes. Two independent transient expression experiments were performed, and plants analyzed for indican at 5 dpi. Bars show average indican accumulation (*n* = 8 infiltrated leaves from individual plants per construct combination) with error bars indicating SD. **(C)** Indole concentrations quantified at 5 dpi in extracts from tobacco plants infiltrated with pBJ1-1 and either pEAQ-EV or pEAQ-AroG*. Bars show averages from a single experiment (*n* = 5 infiltrated leaves from individual plants per construct combination)with error bars indicating standard deviation.

used surrogate assay for CYP2A6 activity that relies on 7-hydroxylation of coumarin into umbelliferone (Kim et al., 2005). This assay revealed that CYP2A_B49 was much more efficiently reduced by AtFd1 (V_{max}/K_M = 13.8) compared to the flavodoxins (Figure 5C, Table 2), with 7 times higher catalytic efficiency than $IsiB^{6803}$ (V_{max}/K_M = 1.9) and 22 times higher catalytic efficiency than $HsCPR^{62-241}$ V_{max}/K_M = 0.6). Surprisingly, despite deriving from the native human CPR and reduced quite efficiently by photosystem I, $HsCPR^{62-241}$ was the worst redox partner for light-driven CYP2A_B49 activity. To go further and investigate how the electron carrier proteins would support CYP2A_B49 activity in the presence of a competing electron sink measured its light-driven activity in the presence and absence of the competing enzyme FNR. FNR readily oxidizes ferredoxins and flavodoxins to reduce $NADP^+$ to NADPH and consumes most of the photosynthetic reducing power in chloroplasts (Holfgrefe et al., 1997; Backhausen et al., 2000). Adding FNR and $NADP^+$ in AtFd1-driven assays decreased CYP2A_B49 activity by about 80% compared with assays without FNR (Figure 5D). CYP2A_B49 activity was mostly unaffected in presence of FNR when flavodoxins were used to supply electron (Figure 5D). Despite the ability of $IsiB^{6803}$ and $HsCPR^{62-241}$ to support CYP2A_B49 activity with equal efficiency whether in the presence of FNR or not, CYP2A_B49 activity was still highest in the presence of AtFd1 even when subjected to competition from FNR.

Discussion

A chloroplast targeted indican biosynthesis pathway is functional in tobacco

In this study we demonstrate expression and targeting of three human P450s to chloroplasts in tobacco (Figure 1) and investigate their light-dependent conversion of indole to the indigo precursor indican. Two of these P450s are natural human P450s, and the third a variant generated by DNA shuffling of three CYP2A isoforms. Functional expression of a mutant of

10.3389/fpls.2022.1049177

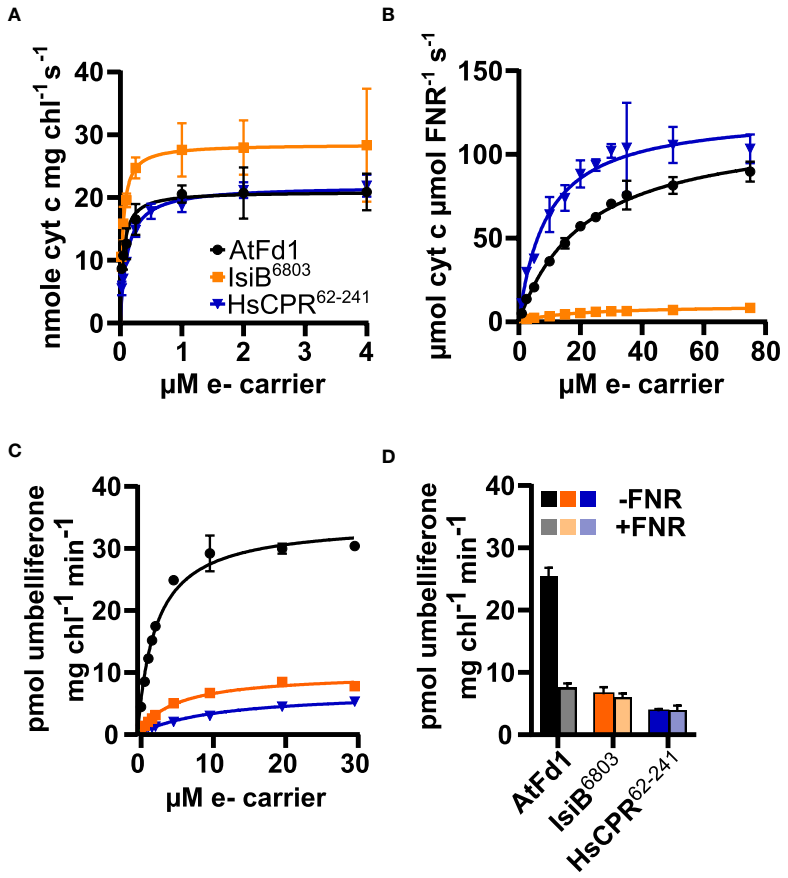


FIGURE 5
Flavodoxin and ferredoxin electron transfer kinetics with photosystem I **(A)**, FNR **(B)** and CYP2A_B49 **(C)**, and CYP2A_B49 + FNR competition assay **(D)**. **(A)** Kinetics of electron transfer from photosystem I to redox partners, measured by redox partner mediated cytochrome *c* reduction using isolated thylakoid membranes supplemented with varying concentrations of redox partners. **(B)** NADPH-dependent reduction of electron carrier proteins by FNR, measured by redox partner mediated cytochrome *c* reduction. **(C)** kinetics of photosynthesis-driven coumarin hydroxylation by CYP2A_B49 in isolated thylakoid membranes in the presence of different redox partners. D, comparison of coumarin hydroxylation rates from CYP2A_B49 with 10 µM redox partner in the presence or absence of 0.6 µM FNR and 1.6 mM $NADP^+$. Plots show mean rates measured from three **(A, D)** or four **(B, C)** technical replicates with error bars indicating ± SD, as well as corresponding Michaelis-Menten fits **(A–C)**.

human CYP2A6 targeted to chloroplast in tobacco was already demonstrated previously (Fräbel et al., 2018). While we did not control for differential accumulation of the three P450s compared, our results were generally in line with previous CYP2A6-CYP2E1 comparisons (Gillam et al., 1999; Warzecha et al., 2007). The present study underscores the applicability of the strategy and shows that photosynthesis can be used widely to drive P450 activity without co-expression of a dedicated

TABLE 2 Average Michaelis-Menten parameters (± SD) of electron transfer reactions between photosystem I, FNR and CYP2A_B49 using the electron carrier proteins tested in this study.

	Photosystem I		FNR		CYP2A_B49	
Electron carrier	V_{max} nmol *cyt c* mg chl^{-1} s^{-1}	K_M µM	V_{max} µmol cyt c µmol FNR^{-1} s^{-1}	K_M µM	V_{max} pmol umbelliferone mg chl^{-1} min^{-1}	K_M µM
AtFd1	21.0 ± 0.8	0.05 ± 0.01	117.9 ± 3.4	21.7 ± 1.5	34.4 ± 0.7	2.5 ± 0.2
$IsiB^{6803}$	28.5 ± 1.3	0.04 ± 0.01	10.0 ± 0.5	17.7 ± 2.3	9.9 ± 0.3	5.2 ± 0.4
$HsCPR^{62-241}$	21.8 ± 0.5	0.11 ± 0.01	125.1 ± 4.9	9.2 ± 1.3	7.0 ± 0.5	11.0 ± 1.7

reductase, because endogenous ferredoxin delivers reducing power to P450 enzymes with very high efficiency (Figure 5, Table 2). Stable expression of CYP2A6 with maize BX1 yielded 0.95 mg g_{DW}^{-1} in 3–4-week-old plants (Warzecha et al., 2007), and was subsequently developed further by introducing bacterial tryptophan halogenases, which allowed synthesis of chloroindican at high yields (0.93 mg g_{DW}^{-1}) in a transient expression system (Fräbel et al., 2018). The titers reported here (~0.2 mg g_{FW}^{-1}) are approximately equal to that of plants stably expressing CYP2A6 (Warzecha et al., 2007) if accounting for the ~80% water content of tobacco leaf (Guo et al., 2019). This shows a remarkable ability of transient expression to produce high yields in shorter time than stable transfection. The speed and scalability of the transient expression system also makes it a valuable tool for fast turnover testing and may in some cases be preferred over stable transgenic plants for biomolecule production (Holtz et al., 2015; Schultz et al., 2019; Kaur et al., 2021). The data also shows that chloroplast targeting of the indican biosynthetic pathway is as viable a strategy as retaining the normal ER targeting of the P450. This is consistent with previous studies, which have shown that eukaryotic P450 enzymes are generally compatible with heterologous thylakoid membrane insertion (Nielsen et al., 2013; Xue et al., 2014; Gangl et al., 2015; Berepiki et al., 2016; Mellor et al., 2016; Li et al., 2019).

Transient indican production in tobacco is limited by CYP2A_B49 activity

We investigate whether by boosting precursor supply or co-expressing redox partners could increase indican productivity. Expressing a feedback-insensitive DAHP synthase (Ding et al., 2014) yielded a nearly 37-fold increase in Trp accumulation and approximately 4-fold increase in indole accumulation (Figure 3, Figure 4C, Table 1). This accumulation is greater than previously reported for *N. tabacum* plants stably expressing feedback insensitive $AroG^{L175Q}$, which yielded 43-, 24- and 10-fold increases in Phe, Tyr and Trp, respectively (Oliva et al., 2020). Despite the increased aromatic amino acid and indole accumulation, we did not observe an increase in indican accumulation when co-expressing AroG* (Figure 4B). We saw a puzzling decrease in indican accumulation when we expressed either AroG* or soluble redox partners from the pEAQ vector alongside the bi-directional pBJ1-1 vector encoding the indican pathway (Figure 4). We could not determine the exact reason for this decrease, but untargeted MS analysis did suggest new or unexplained products (Supplementary Figure 3). Possible explanations include effects due to re-use of the 35S promoter or the CTPs used to target enzymes to chloroplasts, either of which has implications for optimal construct design and should be investigate more thoroughly. We also cannot definitively rule out lower accumulation of CYP2A_B49 as a contributor to this difference. However, previous work has shown that while protein expression can be very different between separate batches of tobacco plants grown at different times (Buyel and Fischer, 2012), protein expression from a given construct within a batch is relatively consistent when infiltrated into leaves of the same age between plants when infiltrated into leaves of the same age between plants (Mellor et al., 2016). Given that these comparisons were always made together on the same batch of plants, and the only difference between infiltrations was the gene inserted in pEAQ vector, we consider reduced CYP2A_B49 expression a less likely explanation. Nevertheless, comparing AroG* and redox partner co-expression indicates that that CYP2A_B49 activity rather than supply of indole or redox equivalents poses the major bottleneck (Figure 4B). This is surprising, as it implies the enzyme is operating near its maximal rate in chloroplasts. The CYP2A_B49 variant used was selected from variant screening based on improved indican productivity in *E. coli*, but its actual kinetic parameters for indole hydroxylation have not been determined. CYP2A6, which performed very similarly in our initial P450 comparison (Figure 1) has k_{cat} and k_{cat}/K_M values for indole hydroxylation of 0.62 s^{-1} and 2.4*10^3 M^{-1} s^{-1}, respectively (Zhang et al., 2009). This is 20 and 40 times less than average k_{cat} and k_{cat}/K_M values surveyed from nearly 2000 enzymes (Bar-Even et al., 2011), and places CYP2A6 in the 16th percentile of surveyed rates. For comparison, *Sorghum* CYP79A1 for which we previously demonstrated photosynthesis-dependent hydroxylation (Mellor et al., 2019) has a k_{cat} of 6.8 s^{-1} (Jensen et al., 2011), in the 42nd percentile and near the average of enzymatic rates surveyed (Bar-Even et al., 2011). Altogether, these lines of evidence point to P450 activity as the major limiting factor for indican productivity in tobacco.

Ferredoxin as a universal P450 redox partner for metabolic engineering

Our *in vitro* comparisons showed that AtFd1 coupled photosynthetic reducing power to CYP2A_B49 activity better than either $IsiB^{6803}$ or $HsCPR^{62-241}$ derived from the human CPR (Figure 5C). Unexpectedly, $HsCPR^{62-241}$ had only 2-fold lower catalytic efficiency as a photosystem I electron acceptor than AtFd1, while $IsiB^{6803}$ had higher efficiency than AtFd1 (Figure 5A, Table 2), which shows that both flavodoxins can integrate with photosynthetic electron transport. Instead, the differences in CYP2A_B49 activity probably arises by faster electro donation from ferredoxin to CYP2A_B49. When we assayed redox partner-dependent CYP2A_B49 activity in the presence of FNR, which competes for reduced ferredoxin and better reflects *in vivo* productivity (Yacoby et al., 2011; Nielsen et al., 2013; Mellor et al., 2016) activity driven by AtFd1 was severely limited (Figure 5D). On the other hand, CYP2A_B49 activity supported by either $IsiB^{6803}$ or $HsCPR^{62-241}$ was

relatively unaffected (Figure 5D). HsCPR$^{62-241}$ has a FMN redox potential of -246 mV at pH 7.5 (Das and Sligar, 2009), and for this redox partner the result is comparable to our previous finding for the flavodoxin-like domain of *Sorghum* CPR2b, whose redox potential of -267 mV (Mellor et al., 2019). Both FMN redox potentials are more positive than the standard redox potential of the $NADP^+$/NADPH couple, the FAD cofactor of FNR (Pueyo et al., 1991) which helps render electron transfer thermodynamically unfavorable. IsiB^{6803} on the other hand possesses a redox potential close to that of ferredoxin by virtue of its role as a ferredoxin surrogate under iron-limiting conditions (Ferreira and Straus, 1994; Pierella Karlusich et al., 2014). As such, there should be no thermodynamic barrier to the passing of reducing power from IsiB^{6803} to FNR in our *in vitro* assays. We instead show that electron transfer between IsiB^{6803} and FNR is slow (Figure 5B), probably due to suboptimal interaction between them, which explains why presence of FNR does not affect IsiB^{6803}-driven CYP2A_B49 activity.

Our data shows that ferredoxin provides optimal coupling of CYP2A_B49 to photosynthetic reducing power. Although coupling P450 activity to photosynthetic reducing power is still a relatively niche field, plant and cyanobacterial ferredoxins has been used as surrogate electron transfer partners in many reconstituted P450 systems (Wirtz et al., 2000; Jackson et al., 2002; Goñi et al., 2009; Jensen et al., 2012; Zhang et al., 2018; Mie et al., 2020; Zhu et al., 2020; Liu et al., 2022). Recently a survey of 16 microbial ferredoxins found ferredoxin from the cyanobacteria *Synechococcus elongatus* PCC 7942 to be the most efficient redox partner for 5 different P450s (Zhang et al., 2018). This was later expanded to include further 2 bacterial P450s favoring non-native cyanobacterial ferredoxin redox partners (Liu et al., 2022). Together with our own findings, these results suggest that plant and cyanobacterial [2Fe-2S] ferredoxins may be ideal generalist P450 redox partners. While this may seem surprising, many ancestral P450 redox partners were likely to have been either ferredoxins or flavodoxins (Paine et al., 2005), which implies the ability to interact with both redox partners could have been present early in the evolutionary history of this class of enzymes and was maintained after the rise of eukaryotes. At the same time ferredoxins have by virtue of their role as central redox distribution hubs been constrained by evolution to retain interactions with hundreds of enzymes (Hanke et al., 2011; Peden et al., 2013; Goss and Hanke, 2014). Such conservation of redox partner interactions in both P450s and ferredoxins explains why electron transfer complementarity was maintained across the evolutionary divide that confined these enzymes to different cellular compartments.

Improving yields from photosynthesis-based indigo production

Overcoming the limitation on indican biosynthesis posed by the activity of CYP2A_B49 in the system described here can be approached either by increasing the expression of the P450 or by screening for variants or enzymes with higher activity. Increased accumulation of indican was obtained when encoding both CYP2A_B49 and tnaA genes on the same vector (Figure 2), but further improvements could be achieved by systematic optimization of construct design. Stably engineering the pathway for expression from the chloroplast genome may be the optimal strategy to reach maximal expression of enzymes for a given biosynthetic pathway in tobacco. Several reports have demonstrated high-level production of small molecules by introducing biosynthetic pathways into the chloroplast genome (Viitanen et al., 2004; Fuentes et al., 2016; Lu et al., 2017), and expression levels of individual proteins expressed from the chloroplast genomes have reached as high as 70% of total leaf protein (Ruhlman et al., 2010). Furthermore, the bacterial-like nature of the plastid genetic machinery has allowed the development of an inducible riboswitch-based T7 promoter system, which is highly advantageous for introduction of pathways whose activity cause deleterious effects on the host plant (Emadpour et al., 2015; Agrawal et al., 2022). Despite being an incredibly powerful approach towards plant-based biomanufacturing, generation of transplastomic plants is difficult and time consuming and far from routine in all but a few academic labs. In this regard, the scalability and ease of transient expression approaches allow them to remain viable alternatives towards high-level small molecule production (Reed et al., 2017).

Only a handful of P450s are currently known to perform indole hydroxylation at high rates. Early indole hydroxylating variants of P450 BM3 had K_M for indole at 2 mM, which meant that overall catalytic efficiency ($1.4*10^3$ M^{-1} s^{-1}) remained lower than those of human P450s (Li et al., 2008). Since, more recent BM3 variants showing improved rates (6.7 s^{-1}) and K_M values (140 μM), and resultant much improved catalytic efficiency of $47.9*10^3$ M^{-1} s^{-1} (Rousseau et al., 2020). A bacterial CYP153 capable of hydroxylating indole was recently found to by screening homologs across bacterial genomes for conversion of indole to indigo (Fiorentini et al., 2018). This enzyme originates from an unclassified *Pseudomonas* accession (sp. 19-rlim) and is a soluble ferredoxin-dependent P450 whose kinetic characteristics have not been described. Both it and P450s BM3 variants may thus be interesting candidates for targeting to chloroplasts.

Several non-P450 enzymes, mostly deriving from prokaryotic indole metabolism, are known to catalyze formation of indigo. Bacterial flavin-containing monooxygenases (FMO) are NADPH-dependent enzymes, some of which have been found to produce indigo from indole (Choi et al., 2003; Ameria et al., 2015), and have been applied successfully in microbial indigo fermentation schemes (Han et al., 2011; Hsu et al., 2018). A flavin-dependent monooxygenase recently identified in the indican-producing plant *Polygonum tinctorium* (Inoue et al., 2021) could be the missing indole hydroxylase in the plant indican biosynthetic pathway. Targeting FMOs to chloroplasts would shift competition for reducing equivalents to NADPH instead of ferredoxin. Improved redox supply could be achieved in this system by engineering FMOs with higher NADPH affinity (Cahn et al., 2017; Jiang et al., 2020) or increasing NADPH supply (Assil-Companioni et al., 2020). Some bacterial naphthalene dioxygenases (NDO) and phenol hydroxylases (PHO) also hydroxylate indole (Fabara and Fraaije, 2020). The NDO are ferredoxin-dependent non-heme iron enzymes, while PHO obtain electrons from NADPH *via* a reductase comprising fused ferredoxin reductase-ferredoxin domains (Leahy et al., 2003) and as such both enzyme types could potentially be integrated directly with photosynthetic electron transport. NDO and PHO are however also heteromultimeric enzyme complexes dependent on multiple cofactors (Fabara and Fraaije, 2020), and their successful chloroplast targeting will require more complicated constructs designs. Establishing indigo production with these types of enzymes in plants warrants study, but their functional targeting and assembly in chloroplasts remains to be established.

This study shows the complexities one encounters when optimizing plant-based metabolic engineering to improve productivity. Though the tools and approaches in plant metabolic engineering grow increasingly sophisticated, the complex genetic and metabolic context found in plant hosts complicates engineering and makes highly parallelized experimental workflows difficult to implement. Systematic studies of genetic and metabolic pathway designs, and studies aimed at increasing the design toolbox would facilitate knowledge-based optimization in the future. In addition, integrating modelling, omics and targeted analytical workflows for accurate assessment of enzyme and metabolite levels, by now relatively routine in microbial metabolic engineering studies (Nielsen et al., 2022) would facilitate identification of pathway bottlenecks and propel green plant-based bioproduction platforms of the future.

Conclusions

We demonstrate chloroplast-based production of indican from tryptophan using photosynthesis-P450 hydroxylation. We show that indican productivity can be improved by combining both tryptophanase and P450 on one T-DNA. Further, we boost tryptophan accumulation by co-expressing feedback insensitive mutant DAHP synthase, which increases indole but not indican yield. Co-expressing additional redox partners also leads to unaltered indican yield. Further *in vitro* investigation shows that ferredoxin provides optimal coupling of photosynthetic electron transport and P450 activity. Our results demonstrate that plant chloroplasts provide a viable membrane and redox environment for animal P450s. Thus, transient expression with chloroplast targeting can be used to generate active P450s that do not require additional redox partners and has applications for metabolic engineering or could be used to elucidate metabolite profiles of human P450s.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

SM, JB, and MP designed experiments. SM, JB, JI, and CC performed experimental work. SM, JB, CC, TL, EG, and MP analyzed the data. SM wrote the manuscript with help and feedback from JB, JI, CC, TL, EG, and MP. All authors contributed to the article and approved the submitted version.

Funding

SM gratefully acknowledges funding from the Novo Nordisk Foundation (Grant NNF18OC0032350). JI is supported by the Novo Nordisk Foundation (Grant NNF17SA0027704). CC acknowledges support from the Danish National Research Foundation (grant99). TL is supported by a Sapere Aude Starting Grant from the Independent Research Fund Denmark (7026-00041B) and a Novo Nordisk Foundation Emerging Investigator Grant (NNF19OC0055356). MP acknowledges funding from the Novo Nordisk Foundation (Grant NNF15OC0016586).

Acknowledgments

The authors thank Lisbeth Mikkelsen for technical assistance. JB acknowledges support from Queensland University of Technology, the CSIRO Synthetic Biology Future Science Platform, and the Australian Research Council Centre of Excellence in Synthetic Biology.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.1049177/full#supplementary-material

References

Agrawal, S., Karcher, D., Ruf, S., Erban, A., Hertle, A. P., Kopka, J., et al. (2022). Riboswitch-mediated inducible expression of an astaxanthin biosynthetic operon in plastids. *Plant Physiol.* 188, 637–652. doi: 10.1093/plphys/kiab428

Ameria, S. P. L., Jung, H. S., Kim, H. S., Han, S. S., Kim, H. S., and Lee, J. H. (2015). Characterization of a flavin-containing monooxygenase from corynebacterium glutamicum and its application to production of indigo and indirubin. *Biotechnol. Lett.* 37, 1637–1644. doi: 10.1007/s10529-015-1824-2

Assil-Companioni, L., Büchsenschütz, H. C., Solymosi, D., Dyczmons-Nowaczyk, N. G., Bauer, K. K. F., Wallner, S., et al. (2020). Engineering of NADPH supply boosts photosynthesis-driven biotransformations. *ACS Catal* 10, 11864–11877. doi: 10.1021/acscatal.0c02601

Backhausen, J. E., Kitzmann, C., Horton, P., and Scheibe, R. (2000). Electron acceptors in isolated intact spinach chloroplasts act hierarchically to prevent over-reduction and competition for electrons. *Photosynth. Res.* 64, 1–13. doi: 10.1023/A:1026523809147

Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., et al. (2011). The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50, 4402–4410. doi: 10.1021/bi2002289

Behrendorff, J. B. Y. H. (2021). Reductive cytochrome P450 reactions and their potential role in bioremediation. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.649273

Behrendorff, J. B. Y. H., Johnston, W. A., and Gillam, E. M. J. (2013). DNA Shuffling of cytochrome P450 enzymes. *Methods Mol. Biol.* 987, 177–188. doi: 10.1007/978-1-62703-321-3_16

Behrendorff, J. B. Y. H., Sandoval-Ibañez, O. A., Sharma, A., and Pribil, M. (2019). Membrane-bound protein scaffolding in diverse hosts using thylakoid protein CURT1A. *ACS Synth. Biol.* 8, 611–620. doi: 10.1021/acssynbio.8b00418

Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika* 93, 491–507. doi: 10.1093/biomet/93.3.491

Berepiki, A., Hitchcock, A., Moore, C. M., and Bibby, T. S. (2016). Tapping the unused potential of photosynthesis with a heterologous electron sink. *ACS Synth. Biol.* 5, 1369–1375. doi: 10.1021/acssynbio.6b00100

Buyel, J. F., and Fischer, R. (2012). Predictive models for transient protein expression in tobacco (Nicotiana tabacum l.) can optimize process time, yield, and downstream costs. *Biotechnol. Bioeng.* 109, 2575–2588. doi: 10.1002/bit.24523

Cahn, J. K. B., Werlang, C. A., Baumschlager, A., Brinkmann-Chen, S., Mayo, S. L., and Arnold, F. H. (2017). A general tool for engineering the NAD/NADP cofactor preference of oxidoreductases. *ACS Synth. Biol.* 6, 326–333. doi: 10.1021/acssynbio.6b00188

Choi, H. S., Kim, J. K., Cho, E. H., Kim, Y. C., Kim, J., and Kim, S. W. (2003). A novel flavin-containing monooxygenase from methylophaga sp strain SK1 and its indigo synthesis in escherichia coli. *Biochem. Biophys. Res. Commun.* 306, 930–936. doi: 10.1016/s0006-291x(03)01087-8

Das, A., and Sligar, S. G. (2009). Modulation of the cytochrome P450 reductase redox potential by the phospholipid bilayer. *Biochemistry* 48, 12104–12112. doi: 10.1021/bi9011435

Ding, R., Liu, L., Chen, X., Cui, Z., Zhang, A., Ren, D., et al. (2014). Introduction of two mutations into AroG increases phenylalanine production in escherichia coli. *Biotechnol. Lett.* 36, 2103–2108. doi: 10.1007/s10529-014-1584-4

Docimo, T., Reichelt, M., Schneider, B., Kai, M., Kunert, G., Gershenzon, J., et al. (2012). The first step in the biosynthesis of cocaine in erythroxylum coca: the characterization of arginine and ornithine decarboxylases. *Plant Mol. Biol.* 78, 599–615. doi: 10.1007/s11103-012-9886-1

Emadpour, M., Karcher, D., and Bock, R. (2015). Boosting riboswitch efficiency by RNA amplification. *Nucleic Acids Res.* 43, e66. doi: 10.1093/nar/gkv165

Estrada, D. F., Laurence, J. S., and Scott, E. E. (2016). Cytochrome P450 17A1 interactions with the FMN domain of its reductase as characterized by NMR. *J. Biol. Chem.* 291, 3990–4003. doi: 10.1074/jbc.M115.677294

Fabara, A. N., and Fraaije, M. W. (2020). An overview of microbial indigo-forming enzymes. *Appl. Microbiol. Biotechnol.* 104, 925–933. doi: 10.1007/s00253-019-10292-5

Ferreira, F., and Straus, N. A. (1994). Iron deprivation in cyanobacteria. *J. Appl. Phycol* 6, 199–210. doi: 10.1007/BF02186073

Fiorentini, F., Hatzl, A.-M., Schmidt, S., Savino, S., Glieder, A., and Mattevi, A. (2018). The extreme structural plasticity in the CYP153 subfamily of P450s directs development of designer hydroxylases. *Biochemistry* 57, 6701–6714. doi: 10.1021/acs.biochem.8b01052

Fräbel, S., Wagner, B., Krischke, M., Schmidts, V., Thiele, C. M., Staniek, A., et al. (2018). Engineering of new-to-nature halogenated indigo precursors in plants. *Metab. Eng.* 46, 20–27. doi: 10.1016/j.ymben.2018.02.003

Fuentes, P., Zhou, F., Erban, A., Karcher, D., Kopka, J., and Bock, R. (2016). A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. *Elife* 5, 1–26. doi: 10.7554/eLife.13664

Gangl, D., Zedler, J. A. Z., Włodarczyk, A., Jensen, P. E., Purton, S., and Robinson, C. (2015). Expression and membrane-targeting of an active plant cytochrome P450 in the chloroplast of the green alga chlamydomonas reinhardtii. *Phytochemistry* 110, 22–28. doi: 10.1016/j.phytochem.2014.12.006

Gillam, E. M., Aguinaldo, A. M., Notley, L. M., Kim, D., Mundkowski, R. G., Volkov, A. A., et al. (1999). Formation of indigo by recombinant mammalian cytochrome P450. *Biochem. Biophys. Res. Commun.* 265, 469–472. doi: 10.1006/bbrc.1999.1702

Gnanasekaran, T., Vavitsas, K., Andersen-Ranberg, J., Nielsen, A. Z., Olsen, C. E., Hamberger, B., et al. (2015). Heterologous expression of the isopimaric acid pathway in nicotiana benthamiana and the effect of n-terminal modifications of the involved cytochrome P450 enzyme. *J. Biol. Eng.* 9, 24. doi: 10.1186/s13036-015-0022-z

Goñi, G., Zöllner, A., Lisurek, M., Velázquez-Campoy, A., Pinto, S., Gómez-Moreno, C., et al. (2009). Cyanobacterial electron carrier proteins as electron

donors to CYP106A2 from bacillus megaterium ATCC 13368. *Biochim. Biophys. Acta (BBA) - Proteins Proteomics* 1794, 1635–1642. doi: 10.1016/j.bbapap.2009.07.012

Goss, T., and Hanke, G. (2014). The end of the line: can ferredoxin and ferredoxin NADP(H) oxidoreductase determine the fate of photosynthetic electrons? *Curr. Protein Pept. Sci.* 15, 385–393. doi: 10.2174/1389203715666140327113733

Guo, G., Li, B., Liu, C., Jin, X., Wang, Z., Ding, M., et al. (2019). Characterization of moisture mobility and diffusion in fresh tobacco leaves during drying by the TG–NMR analysis. *J. Therm. Anal. Calorim* 135, 2419–2427. doi: 10.1007/s10973-018-7312-x

Han, G. H., Bang, S. E., Babu, B. K., Chang, M., Shin, H.-J., and Kim, S. W. (2011). Bio-indigo production in two different fermentation systems using recombinant escherichia coli cells harboring a flavin-containing monooxygenase gene (fmo). *Process Biochem.* 46, 788–791. doi: 10.1016/j.procbio.2010.10.015

Hanke, G. T., Satomi, Y., Shinmura, K., Takao, T., and Hase, T. (2011). A screen for potential ferredoxin electron transfer partners uncovers new, redox dependent interactions. *Biochim. Biophys. Acta (BBA) - Proteins Proteomics* 1814, 366–374. doi: 10.1016/j.bbapap.2010.09.011

Heinig, U., Gutensohn, M., Dudareva, N., and Aharoni, A. (2013). The challenges of cellular compartmentalization in plant metabolic engineering. *Curr. Opin. Biotechnol.* 24, 239–246. doi: 10.1016/j.copbio.2012.11.006

Holfgrefe, S., Backhausen, J. E., Kitzmann, C., and Scheibe, R. (1997). Regulation of steady-state photosynthesis in isolated intact chloroplasts under constant light: responses of carbon fluxes, metabolite pools and enzyme-activation states to changes of electron pressure. *Plant Cell Physiol.* 38, 1207–1216. doi: 10.1093/oxfordjournals.pcp.a029107

Holtz, B. R., Berquist, B. R., Bennett, L. D., Kommineni, V. J. M., Munigunti, R. K., White, E. L., et al. (2015). Commercial-scale biotherapeutics manufacturing facility for plant-made pharmaceuticals. *Plant Biotechnol. J.* 13, 1180–1190. doi: 10.1111/pbi.12469

Hsu, T. M., Welner, D. H., Russ, Z. N., Cervantes, B., Prathuri, L., Adams, P. D., et al. (2018). Employing a glucosyl biochemical protecting group for a sustainable indigo dyeing strategy. *Nat. Publishing Group*, 14 (3), 1–21. doi: 10.1038/nchembio.2552

Huttanus, H. M., and Feng, X. (2017). Compartmentalized metabolic engineering for biochemical and biofuel production. *Biotechnol. J.* 12, 1700052. doi: 10.1002/BIOT.201700052

Inoue, S., Morita, R., and Minami, Y. (2021). An indigo-producing plant, polygonum tinctorium, possesses a flavin-containing monooxygenase capable of oxidizing indole. *Biochem. Biophys. Res. Commun.* 534, 199–205. doi: 10.1016/j.bbrc.2020.11.112

Jackson, C. J., Lamb, D. C., Marczylo, T. H., Warrilow, A. G. S., Manning, N. J., Lowe, D. J., et al. (2002). A novel sterol 14α-demethylase/ferredoxin fusion protein (MCCYP51FX) from methylococcus capsulatus represents a new class of the cytochrome P450 superfamily. *J. Biol. Chem.* 277, 46959–46965. doi: 10.1074/jbc.M203523200

Jander, G., Norris, S. R., Joshi, V., Fraga, M., Rugg, A., Yu, S., et al. (2004). Application of a high-throughput HPLC-MS/MS assay to arabidopsis mutant screening; evidence that threonine aldolase plays a role in seed nutritional quality. *Plant J.* 39, 465–475. doi: 10.1111/j.1365-313X.2004.02140.x

Jensen, K., Jensen, P. E., and Møller, B. L. (2011). Light-driven cytochrome P450 hydroxylations. *ACS Chem. Biol.* 6, 533–539. doi: 10.1021/cb100393j

Jensen, K., Johnston, J. B., Montellano, P. R. O., and Møller, B. L. (2012). Photosystem I from plants as a bacterial cytochrome P450 surrogate electron donor: terminal hydroxylation of branched hydrocarbon chains. *Biotechnol. Lett.* 34, 239–245. doi: 10.1007/s10529-011-0768-4

Jensen, S. B., Thodberg, S., Parween, S., Moses, M. E., Hansen, C. C., Thomsen, J., et al. (2021). Biased cytochrome P450-mediated metabolism *via* small-molecule ligands binding P450 oxidoreductase. *Nat. Commun.* 2021(12), 1 12, 1–14. doi: 10.1038/s41467-021-22562-w

Jiang, H.-W., Chen, Q., Pan, J., Zheng, G.-W., and Xu, J.-H. (2020). Rational engineering of formate dehydrogenase substrate/cofactor affinity for better performance in NADPH regeneration. *Appl. Biochem. Biotechnol.* 192, 530–543. doi: 10.1007/s12010-020-03317-7

Kang, T.-J., Kwon, T.-H., Kim, T.-G., Loc, N.-H., and Yang, M.-S. (2003). Comparing constitutive promoters using CAT activity in transgenic tobacco plants. *Mol. Cells* 16, 117–122. doi: 10.14348/.1970.0.0

Karan, H., Roles, J., Ross, I. L., Ebrahimi, M., Rackemann, D., Rainey, T., et al. (2022). Solar biorefinery concept for sustainable co-production of microalgae-based protein and renewable fuel. *J. Clean Prod* 368, 132981. doi: 10.1016/J.JCLEPRO.2022.132981

Karimi, M., Inzé, D., and Depicker, A. (2002). GATEWAYTM vectors for agrobacterium-mediated plant transformation. *Trends Plant Sci.* 7, 193–195. doi: 10.1016/S1360-1385(02)02251-3

Kaur, M., Manchanda, P., Kalia, A., Ahmed, F. K., Nepovimova, E., Kuca, K., et al. (2021). Agroinfiltration mediated scalable transient gene expression in genome edited crop plants. *Int. J. Mol. Sci.* 22, 10882. doi: 10.3390/ijms221910882

Kim, D., Wu, Z.-L., and Guengerich, F. P. (2005). Analysis of coumarin 7-hydroxylation activity of cytochrome P450 2A6 using random mutagenesis. *J. Biol. Chem.* 280, 40319–40327. doi: 10.1074/jbc.M508171200

Kubota, T., Yoshikawa, S., and Matsubara, H. (1992). Kinetic mechanism of beef heart ubiquinol:cytochrome c oxidoreductase. *J. Biochem.* 111, 91–98. doi: 10.1093/oxfordjournals.jbchem.a123725

Lai, Z., Tsugawa, H., Wohlgemuth, G., Mehta, S., Mueller, M., Zheng, Y., et al. (20172017). Identifying metabolites by integrating metabolome databases with mass spectrometry cheminformatics. *Nat. Methods* 15, 1 15, 53–56. doi: 10.1038/nmeth.4512

Lans, I., Frago, S., and Medina, M. (2012). Understanding the FMN cofactor chemistry within the anabaena flavodoxin environment. *Biochim. Biophys. Acta Bioenerg.* 1817, 2118–2127. doi: 10.1016/j.bbabio.2012.08.008

Laursen, T., Lam, H. Y. M., Sørensen, K. K., Tian, P., Hansen, C. C., Groves, J. T., et al. (2021). Membrane anchoring facilitates colocalization of enzymes in plant cytochrome P450 redox systems. *Commun. Biol.* 2021(4), 1 4, 1–1 4, 9. doi: 10.1038/s42003-021-02604-1

Leahy, J. G., Batchelor, P. J., and Morcomb, S. M. (2003). Evolution of the soluble diiron monooxygenases. *FEMS Microbiol. Rev.* 27, 449–479. doi: 10.1016/S0168-6445(03)00023-8

Lian, L.-Y., Widdowson, P., McLaughlin, L. A., and Paine, M. J. I. (2011). Biochemical comparison of anopheles gambiae and human NADPH P450 reductases reveals different 2′-5′-ADP and FMN binding traits. *PloS One* 6, e20574. doi: 10.1371/journal.pone.0020574

Lichtenthaler, H. K. (1987). "[34] chlorophylls and carotenoids: pigments of photosynthetic biomembranes," in *Methods in enzymology*, (Orlando, FL: Academic Press) 350–382. doi: 10.1016/0076-6879(87)48036-1

Li, H., Mei, L., Urlacher, V. B., and Schmid, R. D. (2008). Cytochrome P450 BM-3 evolved by random and saturation mutagenesis as an effective indole-hydroxylating catalyst. *Appl. Biochem. Biotechnol.* 144, 27–36. doi: 10.1007/s12010-007-8002-5

Li, J., Mutanda, I., Wang, K., Yang, L., Wang, J., and Wang, Y. (2019). Chloroplastic metabolic engineering coupled with isoprenoid pool enhancement for committed taxanes biosynthesis in nicotiana benthamiana. *Nat. Commun.* 2019 (10), 1 10, 1–12. doi: 10.1038/s41467-019-12879-y

Liu, X., Li, F., Sun, T., Guo, J., Zhang, X., Zheng, X., et al. (2022). Three pairs of surrogate redox partners comparison for class I cytochrome P450 enzyme activity reconstitution. *Commun. Biol.* 2022(5), 1 5, 1–1 5,11. doi: 10.1038/s42003-022-03764-4

Liu, X., Zhu, X., Wang, H., Liu, T., Cheng, J., and Jiang, H. (2020). Discovery and modification of cytochrome P450 for plant natural products biosynthesis. *Synth. Syst. Biotechnol.* 5, 187–199. doi: 10.1016/j.synbio.2020.06.008

Lu, Y., Stegemann, S., Agrawal, S., Karcher, D., Ruf, S., and Bock, R. (2017). Horizontal transfer of a synthetic metabolic pathway between plant species. *Curr. Biol.* 27, 3034–3041.e3. doi: 10.1016/j.cub.2017.08.044

Madhavan, A., Arun, K. B., Sindhu, R., Krishnamoorthy, J., Reshmy, R., Sirohi, R., et al. (20212021). Customized yeast cell factories for biopharmaceuticals: from cell engineering to process scale up. *Microbial Cell Factories* 20, 1 20, 1–17. doi: 10.1186/S12934-021-01617-Z

Matsumura, T., Kimata-Ariga, Y., Sakakibara, H., Sugiyama, T., Murata, H., Takao, T., et al. (1999). Complementary DNA cloning and characterization of ferredoxin localized in bundle-sheath cells of maize leaves. *Plant Physiol.* 119, 481–488. doi: 10.1104/pp.119.2.481

Mayhew, S. G., and Massey, V. (1969). Purification and characterization of flavodoxin from peptostreptococcus elsdenii. *J. Biol. Chem.* 244 (4), 794–802. doi: 10.1016/S0021-9258(18)91858-2

Mellor, S. B., Nielsen, A. Z., Burow, M., Motawia, M. S., Jakubauskas, D., Møller, B. L., et al. (2016). Fusion of ferredoxin and cytochrome P450 enables direct light-driven biosynthesis. *ACS Chem. Biol.* 11, 1862–1869. doi: 10.1021/acschembio.6b00190

Mellor, S. B., Vinde, M. H., Nielsen, A. Z., Hanke, G. T., Abdiaziz, K., Roessler, M. M., et al. (2019). Defining optimal electron transfer partners for light-driven cytochrome P450 reactions. *Metab. Eng.* 55, 33–43. doi: 10.1016/j.ymben.2019.05.003

Mie, Y., Yasutake, Y., Takayama, H., and Tamura, T. (2020). Electrochemically boosted cytochrome P450 reaction that efficiently produces 25-hydroxyvitamin D3. *J. Catal* 384, 30–36. doi: 10.1016/j.jcat.2020.02.012

Mirza, N., Crocoll, C., Olsen, C. E., and Halkier, B. A. (2016). Engineering of methionine chain elongation part of glucoraphanin pathway in *E. coli*. *Metabolic Engineering* 35, 31–37. doi: 10.1016/j.ymben.2015.09.012

Nagaya, S., Kawamura, K., Shinmyo, A., and Kato, K. (2010). The HSPterminator of arabidopsis thaliana increases gene expression in plant cells. *Plant Cell Physiol*, 51, 328–332. doi: 10.1093/pcp/pcp188

Nielsen, A. Z., Mellor, S. B., Vavitsas, K., Wlodarczyk, A. J., Gnanasekaran, T., Perestrello Ramos H de Jesus, M., et al. (2016). Extending the biosynthetic repertoires of cyanobacteria and chloroplasts. *Plant J.* 87, 87–102. doi: 10.1111/tpj.13173

Nielsen, J., Tillegreen, C. B., and Petranovic, D. (2022). Innovation trends in industrial biotechnology. *Trends Biotechnol.* 40, 1160–1172. doi: 10.1016/j.tibtech.2022.03.007

Nielsen, A. Z., Ziersen, B., Jensen, K., Lassen, L. M., Olsen, C. E., Møller, B. L., et al. (2013). Redirecting photosynthetic reducing power toward bioactive natural product synthesis. *ACS Synth. Biol.* 2, 308–315. doi: 10.1021/sb300128r

Oliva, M., Guy, A., Galili, G., Dor, E., Schweitzer, R., Amir, R., et al. (2020). Enhanced production of aromatic amino acids in tobacco plants leads to increased phenylpropanoid metabolites and tolerance to stresses. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.604349

Paine, M. J. I., Scrutton, N. S., Munro, A. W., Gutierrez, A., Roberts, G. C. K., and Roland Wolf, C. (2005). Electron transfer partners of cytochrome P450. cytochrome P450: Structure, mechanism, and biochemistry: Third edition Eds. Paul. R. Ortiz de Montellano. New York: Klüwer Academic/Plenum Publishers. 115–148. doi: 10.1007/0-387-27447-2_4

Peden, E. A., Boehm, M., Mulder, D. W., Davis, R., Old, W. M., King, P. W., et al. (2013). Identification of global ferredoxin interaction networks in chlamydomonas reinhardtii. *J. Biol. Chem.* 288, 35192–35209. doi: 10.1074/jbc.M113.483727

Peyret, H., and Lomonossoff, G. P. (2013). The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mol. Biol.* 83, 51–58. doi: 10.1007/s11103-013-0036-1

Pierella Karlusich, J. J., Lodeyro, A. F., and Carrillo, N. (2014). The long goodbye: the rise and fall of flavodoxin during plant evolution. *J. Exp. Bot.* 65, 5161–5178. doi: 10.1093/jxb/eru273

Podust, L. M., and Sherman, D. H. (2012). Diversity of P450 enzymes in the biosynthesis of natural products. *Nat. Prod Rep.* 29, 1251. doi: 10.1039/c2np20020a

Pueyo, J. J., Gomez-Moreno, C., and Mayhew, S. G. (1991). Oxidation-reduction potentials of ferredoxin-NADP+ reductase and flavodoxin from anabaena PCC 7119 and their electrostatic and covalent complexes. *Eur. J. Biochem.* 202, 1065–1071. doi: 10.1111/j.1432-1033.1991.tb16471.x

Reed, J., Stephenson, M. J., Miettinen, K., Brouwer, B., Leveau, A., Brett, P., et al. (2017). A translational synthetic biology platform for rapid access to gram-scale quantities of novel drug-like molecules. *Metab. Eng.* 42, 185–193. doi: 10.1016/j.ymben.2017.06.012

Renault, H., Bassard, J., Hamberger, B., and Werck-Reichhart, D. (2014). Cytochrome P450-mediated metabolic engineering: current progress and future challenges. *Curr. Opin. Plant Biol.* 19, 27–34. doi: 10.1016/j.pbi.2014.03.004

Roles, J., Yarnold, J., Hussey, K., and Hankamer, B. (2021). Techno-economic evaluation of microalgae high-density liquid fuel production at 12 international locations. *Biotechnol. Biofuels* 14, 133. doi: 10.1186/S13068-021-01972-4

Rousseau, O., Ebert, M. C. C. J. C., Quaglia, D., Fendri, A., Parisien, A. H., Besna, J. N., et al. (2020). Indigo formation and rapid NADPH consumption provide robust prediction of raspberry ketone synthesis by engineered cytochrome P450 BM3. *ChemCatChem* 12, 837–845. doi: 10.1002/cctc.201901974

Ruhlman, T., Verma, D., Samson, N., and Daniell, H. (2010). The role of heterologous chloroplast sequence elements in transgene integration and expression. *Plant Physiol.* 152, 2088–2104. doi: 10.1104/pp.109.152017

Sørensen, M., Andersen-Ranberg, J., Hankamer, B., and Møller, B. L. (2022). Circular biomanufacturing through harvesting solar energy and CO2. *Trends Plant Sci.* 27, 655–673. doi: 10.1016/j.tplants.2022.03.001

Schultz, B. J., Kim, S. Y., Lau, W., and Sattely, E. S. (2019). Total biosynthesis for milligram-scale production of etoposide intermediates in a plant chassis. *J. Am. Chem. Soc.* 141, 19231–19235. doi: 10.1021/jacs.9b10717

Strohmaier, S. J., Huang, W., Baek, J. M., Hunter, D. J. B., and Gillam, E. M. J. (2019). Rational evolution of the cofactor-binding site of cytochrome P450 reductase yields variants with increased activity towards specific cytochrome P450 enzymes. *FEBS J.* 286, 4473–4493. doi: 10.1111/FEBS.14982

Viitanen, P., Devine, A. L., Khan, M. S., Deuel, D. L., van Dyk, D. E., and Daniell, H. (2004). Metabolic engineering of the chloroplast genome using the escherichia coli ubiC gene reveals that chorismate is a readily abundant plant precursor for p-hydroxybenzoic acid biosynthesis. *Plant Physiol.* 136, 4048–4060. doi: 10.1104/pp.104.050054

Warzecha, H., Frank, A., Peer, M., Gillam, E. M. J., Guengerich, F. P., and Unger, M. (2007). Formation of the indigo precursor indican in genetically engineered tobacco plants and cell cultures. *Plant Biotechnol. J.* 5, 185–191. doi: 10.1111/j.1467-7652.2006.00231.x

Wirtz, M., Klucik, J., and Rivera, M. (2000). Ferredoxin-mediated electrocatalytic dehalogenation of haloalkanes by cytochrome P450 cam. *J. Am. Chem. Soc.* 122, 1047–1056. doi: 10.1021/ja993648o

Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K., and Bornscheuer, U. T. (2021). Biocatalysis: enzymatic synthesis for industrial applications. Angewandte Chemie International edition 60, 88–119. doi: 10.1002/anie.202006648

Xue, Y., Zhang, Y., Grace, S., and He, Q. (2014). Functional expression of an arabidopsis p450 enzyme, p-coumarate-3-hydroxylase, in the cyanobacterium synechocystis PCC 6803 for the biosynthesis of caffeic acid. *J. Appl. Phycol* 26, 219–226. doi: 10.1007/s10811-013-0113-5

Yacoby, I., Pochekailov, S., Toporik, H., Ghirardi, M. L., King, P. W., and Zhang, S. (2011). Photosynthetic electron partitioning between [FeFe]-hydrogenase and ferredoxin:NADP+-oxidoreductase (FNR) enzymes *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9396–9401. doi: 10.1073/pnas.1103659108

Zhang, W., Du, L., Li, F., Zhang, X., Qu, Z., Han, L., et al. (2018). Mechanistic insights into interactions between bacterial class I P450 enzymes and redox partners. *ACS Catal* 8, 9992–10003. doi: 10.1021/acscatal.8b02913

Zhang, Z.-G., Liu, Y., Guengerich, F. P., Matse, J. H., Chen, J., and Wu, Z.-L. (2009). Identification of amino acid residues involved in 4-chloroindole 3-hydroxylation by cytochrome P450 2A6 using screening of random libraries. *J. Biotechnol.* 139, 12–18. doi: 10.1016/j.jbiotec.2008.09.010

Zhu, Y., Zhang, Q., Fang, C., Zhang, Y., Ma, L., Liu, Z., et al. (2020). Refactoring the concise biosynthetic pathway of cyanogramide unveils spirooxindole formation catalyzed by a P450 enzyme. *Angewandte Chemie* 132, 14169–14173. doi: 10.1002/ange.202004978

frontiers | Frontiers in Plant Science

TYPE Review
PUBLISHED 06 February 2023
DOI 10.3389/fpls.2023.1116812

OPEN ACCESS

EDITED BY
Xiao-Hong Yu,
Brookhaven National Laboratory (DOE), United States

REVIEWED BY
Jean-David Rochaix,
University of Geneva, Switzerland
Ben M. Long,
Australian National University, Australia

*CORRESPONDENCE
Deserah D. Strand
strandd1@msu.edu

SPECIALTY SECTION
This article was submitted to Plant Systems and Synthetic Biology, a section of the journal Frontiers in Plant Science

RECEIVED 05 December 2022
ACCEPTED 23 January 2023
PUBLISHED 06 February 2023

CITATION
Strand DD and Walker BJ (2023) Energetic considerations for engineering novel biochemistries in photosynthetic organisms.
Front. Plant Sci. 14:1116812.
doi: 10.3389/fpls.2023.1116812

Energetic considerations for engineering novel biochemistries in photosynthetic organisms

Deserah D. Strand[1]* and Berkley J. Walker[1,2]

[1]U. S. Department of Energy (DOE) Plant Research Laboratory, Michigan State University, East Lansing, MI, United States, [2]Department of Plant Biology, Michigan State University, East Lansing, MI, United States

Humans have been harnessing biology to make valuable compounds for generations. From beer and biofuels to pharmaceuticals, biology provides an efficient alternative to industrial processes. With the continuing advancement of molecular tools to genetically modify organisms, biotechnology is poised to solve urgent global problems related to environment, increasing population, and public health. However, the light dependent reactions of photosynthesis are constrained to produce a fixed stoichiometry of ATP and reducing equivalents that may not match the newly introduced synthetic metabolism, leading to inefficiency or damage. While photosynthetic organisms have evolved several ways to modify the ATP/NADPH output from their thylakoid electron transport chain, it is unknown if the native energy balancing mechanisms grant enough flexibility to match the demands of the synthetic metabolism. In this review we discuss the role of photosynthesis in the biotech industry, and the energetic considerations of using photosynthesis to power synthetic biology.

KEYWORDS

photosynthesis, chloroplast bioenergetics, synthetic biology, chloroplast biotechnology, bioenergetics

Introduction

Plants and algae are attractive platforms for bioengineering pathways for valuable metabolic products because they use freely available solar energy to sustainably power synthetic biology. To make best use of this energy, the localization of engineered synthetic metabolism should be carefully considered relative to where the metabolic precursors are

Abbreviations: ATP, adenosine triphosphate; cyt *bf*, cytochrome *bf* complex; CO_2, carbon dioxide; CBB, Calvin, Benson,Bassham cycle; CF_o, membrane fraction of ATP synthase; CF_1, soluble fraction of ATP synthase; ETC, electron transport chain; FBPase, fructose 1,6 bisphosphatase; NADPH, nicotinamide adenine dinucleotide phosphate; O_2, oxygen; O_2^- superoxide; PCON, photosynthetic control; PG, 2-phosphoglycolate; PGA, 3-phosphoglycerate; PQ, plastoquinone; PQH_2, plastoquinol; PR, photorespiration; PSI, photosystem I; PSII, photosystem II; q_E, exciton quenching; ROS, reactive oxygen species; rubisco, ribulose 1,5 bisphosphate carboxylase/oxygenase; RUBP, ribulose 1,5 bisphosphate; Δp, protonmotive force; ΔpH, fraction of Δp stored as pH; $\Delta\psi$, membrane potential since the light reactions.

produced. In the cytosol and other non-energetic subcellular compartments, the metabolic precursors ATP and reducing equivalents (e.g., NADPH and ferredoxin) required for synthetic metabolism must be supplied by other metabolism or transported from the energetic compartments (i.e., mitochondria or chloroplast). For synthetic metabolism with large demands of ATP and/or reducing equivalents, localization to the chloroplast allows the synthetic metabolism direct access to ATP and reducing equivalents generated by the light dependent reactions in the chloroplast. Traditional bioengineering approaches usually focus on engineering carbon flux but seldom think about the types of energy fluxes required to power synthetic metabolism. In point of fact, one should not consider the light reactions as an independent supplier of ATP or NADPH because their production is intrinsically linked. The light dependent reactions in the thylakoid electron transport chain are constrained to produce a fixed stoichiometry of ATP and reducing equivalents that may not match the newly introduced synthetic metabolism, leading to inefficiency or damage. While photosynthetic organisms have evolved several ways to modify the ATP/NADPH output from their thylakoid electron transport chain, it is unknown if the native energy balancing mechanisms grant enough flexibility to match the demands of the synthetic metabolism. In other words, is native photosynthetic energy production able to support maximal yields of bioproduction from engineered photosynthetic organisms, or will photosynthesis itself need to be re-engineered?

In this review, we discuss bioengineering from an energetics perspective and where this may present a hurdle for using photosynthetic organisms for producing bioproducts of interest.

The chloroplast as a factory

The chloroplast sets photosynthetic eukaryotes apart from other organisms for bioengineering. The chloroplast is the solar cell of the plant, utilizing light energy to generate ATP and reducing power for use in downstream metabolism. This makes these organisms attractive as a bioengineering platform for synthetic metabolism for valuable products because growth and yield is sustained by light; a cheap, abundantly available natural resource. Additionally, the chloroplast itself is the site of many metabolic precursors that can be harnessed for synthetic metabolism (e.g., ATP, reducing equivalents, sugar phosphates, fatty acids, isoprenoids, and amino acids). Enzymes in a desired synthetic pathway that utilize these precursors may be expressed in the nucleus of eukaryotes and targeted to the chloroplast, or integrated into the chloroplast genome, with both approaches having their place in bioengineering.

There are tradeoffs that must be considered when choosing to express transgenes from either the nucleus or the chloroplast. Most nuclear transformation techniques result in random transgene integration and a variety of transgene expression levels. Chloroplast transformation, however, allows for directed integration of transgenes and every chloroplast transformant is genetically identical, resulting in reproducible expression (Bock, 2001; Bock, 2015; Ahmad et al., 2016). Nuclear transformation may lead to undesirable transgene silencing in later generations (De Wilde et al., 2000). While there are approaches to avoid silencing and increase transgene expression in the nucleus, chloroplast transgenes are not subject to silencing, resulting in continued high expression generations later. Therefore, chloroplast transformation may be an approach to avoid the silencing problem completely if the transgene product is ultimately desired in the chloroplast (Bock, 2001).

There are vastly more species with transformable nuclei than chloroplasts, but the number of valuable species amenable to chloroplast transformation is constantly increasing, and established protocols exist for many model organisms and economically important crops (e.g., soybean) (Boynton et al., 1988; Svab and Maliga, 1993; Ahmad et al., 2016; Nimmo et al., 2019; Kaushal et al., 2020; Ruf et al., 2021). Additionally, the molecular toolbox for control of chloroplast transgene expression is continually growing, offering synthetic biologists a variety of options to fine-tune expression levels and timing (Bock, 2022). These molecular transformation tools allow us to utilize the chloroplast as a molecular factory, using the raw materials of the chloroplast to supply the production of valuable molecules. However, synthetic metabolism will be in competition for energy and carbon flux within the native chloroplast metabolism. Strategies for optimizing efficiency and yield of our synthetic metabolism will need to account for the demands of the chloroplast's native metabolism, and how those demands are currently supplied.

The native chloroplast energy demand

Natively, the largest sinks for the ATP and reducing equivalents in the chloroplast are the Calvin-Benson-Bassham (CBB) cycle and photorespiration (Noctor and Foyer, 1998; Walker et al., 2020). The CBB cycle begins when ribulose 1,5 bisphosphate carboxylase/oxygenase (rubisco) catalyzes reaction with CO_2 (carboxylation reaction) and ribulose 1,5 bisphosphate (RuBP), forming two 3-carbon molecules of 3-phosphoglycerate (PGA). PGA is converted to 1,3-bisphosphoglycerate and then to glyceraldehyde-3-phosphate, a portion of which is used in the construction of glucose and other sugars with the remainder feeding back into the regeneration of RuBP, consuming ATP and NADPH in a stoichiometry of 3:2. Alternatively, when rubisco catalyzes reaction with O_2 (oxygenase reaction), one molecule of PGA and one molecule of 2-phosphoglycolate (PG) are formed (Ogren, 1984; Foyer et al., 2009). The PG is inhibitory and must be detoxified in a series of enzymatic reactions that occur across three organelles in a process termed photorespiration (PR) which loses CO_2 in the process. Energetically, the whole-cell demand ratio for ATP : NADPH for photorespiration is 1.75 (Edwards and Walker, 2003; Walker et al., 2020). Regardless of the assumptions used for determining the whole-cell energy demand of photorespiration, the conversion of glycerate to PGA requires an ATP within the chloroplast. This increased consumption in the chloroplast means that in the absence of ATP or NADPH transport, photorespiration increases the relative ATP demand within the chloroplast above that of the CBB cycle.

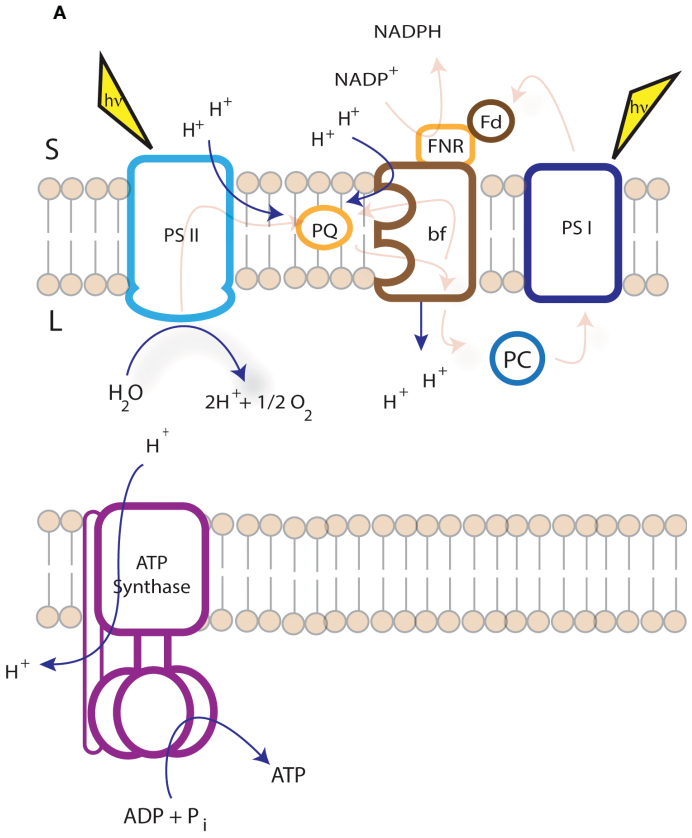
Other pathways in the chloroplast have higher relative reductant demands than the CBB cycle and PR. Both synthesis of fatty acids and desaturation of lipids require reducing equivalents in the form of NADPH, making the demand for reducing equivalents from lipid biosynthesis larger that of the CBB cycle (Ohlrogge and Browse, 1995). Additionally, the methylerythritol 4-phosphate portion of the

isoprenoid biosynthesis pathway in the chloroplast also has a larger demand for reducing equivalents relative to ATP (Zhao et al., 2013). While under normal conditions these metabolic fluxes are much smaller than that of the CBB cycle and PR, metabolic engineering with the goal of increasing flux through these pathways would increase the relative reductant demand within the chloroplast.

The chloroplast energy supply

Since the light reactions of photosynthesis provide all the energy for native and synthetic metabolism in autotrophs, it is essential to understand how light energy is converted to usable chemical energy and what constrains production of this chemical energy. It is also important to understand the constraints to this energy production for optimal bioengineering of reliant synthetic metabolism.

The light reactions of photosynthesis use light excitation to drive a series of redox reactions along an electron transport chain (ETC) to shuttle protons and electrons across the thylakoid membrane generating ATP and reducing power for downstream metabolism [Figure 1, (Hill and Bendall, 1960; Arnon, 1971; Mitchell, 1975; Blankenship, 2002; Mitchell, 2011)]. The coupling of these proton and electron circuits determine the ATP/NADPH produced from the system. When light hits the photosystems, the excited reaction centers



B

ETC: ~1.3 ATP/NADPH
CBB: ↑ ATP/NADPH
PR: ↑ ATP/NADPH
FA: ↓ ATP/NADPH
Protein: ↑ ATP/NADPH

FIGURE 1
Supply and demand of NADPH and ATP within the chloroplast. The proton to electron stoichiometry of proton coupled electron transfer of the light reactions of photosynthesis (as pictured in panel **A**) is fixed at $3H^+/e^-$ and, together with the c-ring structure of the ATP synthase, constrains the electron transport chain production to ~1.3 ATP/NADPH. The light reactions supply multiple metabolic sinks within the chloroplast that do not match the ATP/NADPH resulting in an ATP deficit or surplus. **(B)** indicates pathways of interest for metabolic engineering, and their difference in demand (within the chloroplast) from the ATP/NADPH production of the electron transport chain. CBB, Calvin-Benson-Bassham cycle; PR, photorespiration; FA, Fatty acid biosynthesis; Protein, Protein synthesis.

lose an electron. At photosystem II (PSII), electron loss at P680 supplies the oxidative force ($P680^+$) to release electrons from water, reducing $P680^+$ and releasing protons into the lumen. Electron transfer falls downhill in energy to plastoquinone (PQ), the reduction of which leads to the uptake of protons from the stroma. These protons are released into the lumen when plastoquinol (PQH_2) is oxidized at the cytochrome *bf* (cyt *bf*) complex. Electron transfer continues through plastocyanin to reduce the oxidized photosystem I (PSI) reaction center, $P700^+$. When light energy oxidizes P700, the electron travels through PSI to the soluble electron carrier ferredoxin which can provide reducing power to metabolism, redox regulation, or be used to reduce $NADP^+$ to NADPH. NADPH is used as a two-electron donor to a wide variety of metabolism in the chloroplast as discussed above.

The proton-coupled electron transfer reactions at PSII and cyt *bf* lead to proton accumulation in the thylakoid lumen, generating an electrochemical proton gradient used for ATP synthesis at the ATP synthase. At PSII, water splitting results in the release of one proton into the lumen per electron entering the electron transport chain. When PQH_2 is oxidized at the cyt *bf* complex, electrons enter a bifurcated pathway, one of which is eventually passed to PSI, and another is passed to another site on cyt *bf* to half reduce another PQ [i.e., the Q-cycle, (Mitchell, 1975)]. Therefore, for each electron passed from cyt *bf* to PSI, two protons are deposited into the lumen. Across the membrane, the difference in charge imposed by this proton gradient comprises a membrane potential ($\Delta\psi$) and the difference in pH generates a pH gradient (ΔpH) (Cruz et al., 2001). Together, these comprise the protonmotive force (Δp) for ATP synthesis *via* the ATP synthase (Nicholls and Ferguson, 2002; Kramer et al., 2003). The chloroplast ATP synthase is comprised of two complexes, the CF_o in the membrane, and the CF_1 in the stroma (Hahn et al., 2018). The CF_o contains a 14-subunit c-ring and a proton half channel, releasing 14 protons per one full turn (Seelert et al., 2003). This turn moves the g-subunit, which in turn drives the conformational change in the CF_1 domain, comprised of 3 heterodimers, and responsible for ATP synthesis. Therefore, 14 protons are required to synthesize 3 ATP, and $NADP^+$ reduction to NADPH requires 2 electrons from the electron transport chain.

From the H^+/e^- at PSII and cyt *bf*, there is a total stoichiometry of the ETC of $3H^+/e^-$, which results in an ATP/NADPH of ~1.3, in a perfectly coupled system. This means that, the largest sinks for ATP and NADPH in the chloroplast (i.e., CBB cycle and photorespiration) are running at an ATP deficit. The ATP/NADPH production stoichiometry is flexible to a certain extent by the activation of, or leak of electrons into, alternative electron transfer pathways [discussed in depth elsewhere, e.g., (Kramer and Evans, 2011; Strand et al., 2016; Alric and Johnson, 2017; Walker et al., 2020)]. The most studied alternative pathways lead to an increase in the relative production of ATP from the ETC. This is likely due to the largest sinks in the chloroplast imposing an ATP deficit on the ETC (discussed below), however, with the introduction of synthetic biology into the chloroplast, we must consider the impact of new or altered metabolic sinks on chloroplast energetics.

Consequences of energetic imbalance

To optimize efficiency and avoid damage, the chloroplast must balance the supply and demand of ATP and reducing equivalents. Energy demand from native metabolism in the chloroplast does not match the calculated output of the light reactions, therefore it is likely that throughout the organism's life cycle it experiences periods of ATP deficits or surplus. What problem do these conditions pose for the plant?

When ATP demand relative to NADPH exceeds the supply of the light reactions (ATP deficit, Figure 2A), sinks for reducing power may be closed as they do not have ATP as substrate to turnover acceptors. For example, if ATP is limiting to the CBB cycle, 1,3-bisphosphoglycerate is not produced from 3-phosphoglycerate. Since 1,3-bisphosphoglycerate is an electron acceptor, $NADP^+$ is not being regenerated for reduction by PSI. This shifts the NADPH/$NADP^+$ to a more reduced state, lowering the redox potential of the acceptor side of PSI. A more reduced acceptor side of PSI is dangerous because it will increase the likelihood of electrons being passed to oxygen. Reduction of oxygen generates superoxide ($O_2^{\cdot-}$), a reactive oxygen species that can damage protein and lipid, and must be detoxified [i.e., through the water-water cycle, (Asada, 2000)]. This detoxification itself diverts electrons from NADPH, and activates another electron sink (i.e, cyclic electron flow) to increase ATP/NADPH production (Casano et al., 2001; Strand et al., 2015).

However, if a large enough ATP sink was introduced into the chloroplast, it is not clear if the self-balancing machinery of the chloroplast is robust enough to meet the new demand without side effects like decreased growth. An example of when the self-balancing machinery fails to meet a new metabolic demand may be seen in a mutant of the chloroplast fructose 1,6 bisphosphatase (FBPase) (Livingston et al., 2010). This mutant seemingly survives disruption to the main metabolic sink in the chloroplast by bypassing the FBPase lesion though the glucose-6-phosphate shunt, which has increased relative ATP demand (Sharkey and Weise, 2016). Despite activation of multiple alternative electron transfer pathways to increase ATP supply from the ETC (Livingston et al., 2010; Strand et al., 2015), these plants grow slower and smaller.

This may, in part, be due to the secondary impact cyclic electron flow around PSI (CEF), an alternative electron transport pathway that is increased in the FBPase knockout which may decrease the efficiency of photosynthesis. CEF increases proton translocation into the lumen relative to electron transfer from PSII. Additional proton translocation can increase the Δp in the lumen, activating pH dependent feedback regulation i.e., downregulation of light harvesting (exciton quenching, q_E) and electron transfer at cyt *bf* (photosynthetic control, PCON). In other words, the responses used to mitigate the ATP deficit may lead to total decrease in output of the light reactions. In an industrial process, this would be an undesired side effect limiting the yield of the engineered pathways.

While the chloroplast has many identified routes of alternative electron transfer to deal with an ATP deficit, much less in known about how an ATP surplus (or, alternatively, a reducing equivalent deficit). For those routes that are proposed, there is little evidence that they contribute substantially to chloroplast energy balancing in the light (Flügge et al., 2011; Strand et al., 2017). The prevalence of pathways to mitigate an ATP deficit rather than mitigate an ATP surplus may be due to selective pressure stemming from native metabolism being poised towards an ATP deficit as discussed above (Walker et al., 2020). However, many proposed synthetic biology pathways result in an ATP surplus such as increased lipid production

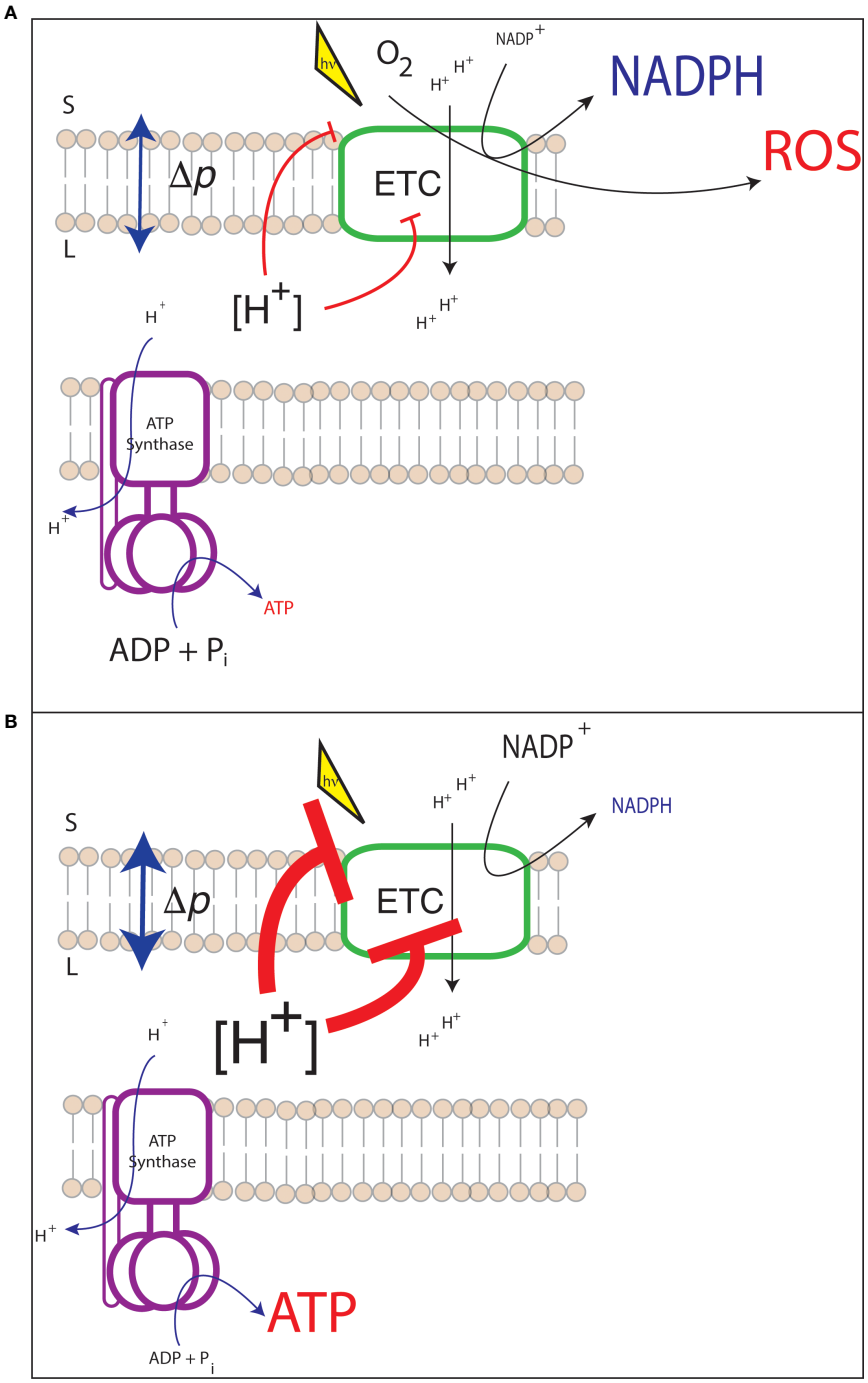


FIGURE 2
Possible consequences of energetic mismatch between the electron transport chain (ETC) and downstream metabolism. **(A)** An ATP deficit may result in the closure of the acceptor side of PSI and the generation of harmful reactive oxygen species (ROS); **(B)** An ATP surplus may lead to substrate limitation of ATP synthase and the downregulation of light harvesting and electron transfer resulting in lower photosynthetic efficiency.

in microalgae and photorespiratory bypasses in plants (Miller et al., 2020).

Under an ATP surplus condition the threat of damage to the chloroplast is not a primary concern, instead the efficiency of the light reactions will decrease. This inefficiency would occur through a shift in the [ATP]/[ADP] [P_i] pool which may lead to an increase in Δp required to drive ATP synthesis. All else being equal, increased Δp leads to an activation of q_E and PCON (Figure 2B), decreasing light harvesting and electron transfer rates. In the case of a synthetic pathway with an increased electron demand, activation of these photoprotective processes would lead to decreased yield of the synthetic pathway. This response to increased electron demand may be a limiting factor for any synthetic pathway with high reductant demand, such as increased lipid production.

The conditions described above are an important consideration for bioengineering pathways relying on output from the light reactions of photosynthesis. This is because final product yield may ultimately be limited by the light reactions if new metabolism exceeds the innate

energetic flexibility of the chloroplast, even if the engineered metabolism yields the expected product to some degree. To increase product yield, it may be enticing to modify the regulatory processes that are activated in response to energetic imbalance. However, feedback regulation that limits light utilization and electron transfer (i.e., q_E and PCON) are important to protect the photosystems from photodamage and in the case of PCON, it is *essential* for plant survival under fluctuating conditions (Müller et al., 2001; Suorsa et al., 2012). This is an important point: energetic limitations can't simply be rectified by eliminating feedback regulation, and instead may require additional engineering to optimize our desired yield.

Case studies in chloroplast energetics

Since the invention of plant transformation, there have been many successful applications of bioengineering metabolism to produce valuable metabolites, either by the introduction of synthetic metabolism or altering the flux through metabolic pathways to favor a specific molecule [reviewed in detail in (O'Neill and Kelly, 2017)]. While there have been many successes without energetic considerations, it may be worthwhile to consider the ATP and NADPH requirements when yield from synthetic pathways is lower than expected. Here we offer examples of synthetic metabolism that, if successfully implemented into a photosynthetic organism would drastically alter the energetic demands in the chloroplast and may need additional engineering to ensure maximal yield of the desired product.

Due to the loss of CO_2 during the detoxification and recycling of PG, photorespiration is thought to be a large inefficiency in C3 plants and is an active area of research for improvement of crop productivity (Fu et al., 2022b). There have been several approaches to decrease the rate of photorespiration, e.g., by introducing a bypass modifying the compartment of CO_2 release (Peterhansel et al., 2010; South et al., 2019), or compartmentalization of rubisco (Lin et al., 2014). However, an alternative approach to decreasing biomass loss due to photorespiration is the complete bypass of rubisco to fix carbon. One approach that has been shown to operate *in vitro* is the so called 'Cetch' cycle (Schwander et al., 2016; Miller et al., 2020), a completely synthetic pathway for CO_2 fixation achieved by combining enzymes from multiple organisms across the tree of life. This approach is interesting due the pathway's increased relative demand in reducing equivalents. The challenges of introducing this cycle into a plant are extensive (e.g., expression, availability of cofactors), but the prospect does introduce an intriguing case study of the need for multiple engineering targets to optimize yield. Since the demand of this cycle for reductant is extremely altered (0.66 ATP/NADPH) from that of native CO_2 fixation (1.5 ATP/NADPH), organisms utilizing the 'Cetch' cycle would certainly require secondary bioengineering to overcome the substantial change in chloroplast energetics for growth to be sustainable.

Another engineering target is in the use of photosynthetic organisms to produce biofuel feedstocks (Blatti et al., 2013). One attractive approach to sustainable biofuel production is the use of lipid rich algae. There has been substantial research effort into both identifying algal species that are suitable for robust lipid production and developing molecular toolkits to allow for their genetic modification (Alishah Aratboni et al., 2019). While there have been some successes in increasing lipid production or shifting lipid production to more valuable lipid species, more effort is needed to make algae an economically viable biofuel feedstock (Blatti et al., 2013; Alishah Aratboni et al., 2019; Gilmour, 2019; Verma and Kuila, 2020). From an energetic (i.e., ATP : NADPH budget) perspective, increasing relative lipid production has different considerations than increasing overall biomass. The latter may be approached by looking at metabolic bottlenecks, but substantial flux through the fatty acid biosynthetic pathway, with its relatively high NADPH demand, may introduce a bottleneck to the light reactions (Figure 2B).

Due to the targeted nature of synthetic biology, the energetic demands of the synthetic pathway can be determined from its ATP and reductant demands. This presents an opportunity to incorporate the new metabolism into existing models [e.g., (Noctor and Foyer, 1998)] and estimate the impact the synthetic pathways may have on cellular energetics. Several flux models exist quantifying metabolic flux under various environmental conditions (Ma et al., 2014; Wieloch and Sharkey, 2022; Fu et al., 2022a) and may help to gauge the impact introduction of new metabolic sinks may have on energetics relative to native metabolism. This could also serve as a starting point for the design of accessory metabolism to offset any energetic imbalances imposed by the introduced metabolism.

Translation to crops

Broadly, increasing photosynthetic output by targeting inefficiencies has been shown to increase overall plant biomass [e.g.(Driever et al., 2017; South et al., 2019; Chen et al., 2020; Wang et al., 2020)]. In biofuel crops where general biomass is utilized for fermentation, this will likely result in the increased yield of desired product, but most bioengineering approaches are concerned with increasing a specific plant product (e.g., grain, lipids). Due to the complex regulation of carbon partitioning in crops (Yu et al., 2015; Paul, 2021), it is unclear if targeting photosynthetic inefficiencies or energetics will translate to increased grain or fruit production.

Despite these potential issues, engineering approaches that increase CO_2 assimilation have yielded higher overall biomass, including an increase in grain (Driever et al., 2017). Since many breeding successes have been measured by increasing yield while decreasing overall plant size [i.e., increased harvest index (HAY, 1995)], increasing grain yield by increasing overall crop biomass may increase the land use need, offsetting our perceived increases.

Improving photosynthesis for growth and resilience

Despite the uncertainty of downstream partitioning of biomass, photosynthesis researchers have a goal focused on improving the efficiency of photosynthesis (Long et al., 2006; Ort et al., 2015). With increased understanding of the regulation of these processes, we may increase the amount of light energy harnessed into driving CO_2 assimilation. Regardless of the downstream success in biomass partitioning, these increases will be an important building block in our steps to improve crops. There have been several approaches taken to improving the electron transport chain, many of which have resulted in small increases photosynthetic efficiency and/or in downstream carbon assimilation or overall plant biomass.

One large target of improvement is the group of processes activated by the accumulation of the ΔpH component of Δ*p* (i.e., q^E and PCON). Even though q_E and PCON are important to protect the thylakoid electron transport chain from damage, they are often cited as points for crop improvement to minimize energy loss and increase electron transfer rates (Murchie and Niyogi, 2011). In the case of an ATP surplus introduced by synthetic biology, these processes may also be a target to increase electron transfer to synthetic electron sinks. This should be undertaken with caution, as it may seem attractive to alter the sensitivity of these photoprotective processes to increase electron transfer, but this approach will also open our organisms to increased risk of photodamage. Under controlled environmental conditions, this may not be problematic, however, if engineered plants are grown in greenhouse or field conditions, they will experience rapidly changing light conditions. Without feedback regulation excess excitation energy can lead to the damage of PSII which must be repaired, leading to the long-term downregulation of light harvesting (Müller et al., 2001). Additionally, under high or rapidly fluctuating light, unregulated transfer of electrons through cyt *bf* can lead to damage of PSI, a potentially lethal situation for the plant (Suorsa et al., 2012).

Despite these concerns, there has been limited success with approaches to increase the flexibility of feedback regulation. Overexpression of cyt *bf* components (the rate limiting step of the electron transport chain) in a C4 plant has led to small increases in electron transfer and CO_2 assimilation (Ermakova et al., 2019). This suggests there is some limitation due to protein content of the cyt *bf* complex, and that subtle increases of electron transfer through *bf* may be a route to improving photosynthetic efficiency. This is especially of interest because overexpression of cyt *bf* would not impact dynamic PCON, maintaining PSI photoprotection. Increased expression of a K^+/H^+ antiporter involved in exchanging ΔpH for Δψ during fluctuating light changes increased photosynthetic efficiency due to the increased capacity to relax q_E (Armbruster et al., 2016). Overexpression of q_E components, which has been shown to alter the sensitivity of q_E to the ΔpH component of Δ*p* (Li et al., 2002), has led to biomass increases in tobacco and soybean (Kromdijk et al., 2016; De Souza et al., 2022). However, in Arabidopsis the results were opposite (Garcia-Molina and Leister, 2020). Interpretation of these conflicting results is complicated, as the xanthophyll cycle component of NPQ is intrinsically linked to abscisic acid biosynthesis, the plant hormone involved in plant water relations (Niyogi et al., 1998).

While the light reactions supply ATP and NADPH for downstream carbon assimilation, except for low light, they do not appear to be the major limitation on bulk photosynthetic yield. There have also been improvements in assimilation and growth increases when the carbon reactions of photosynthesis are targeted, without modification of the light reactions. For example, overexpression of sebdoheptulose 1,7 bisphosphatase, the rate limiting step in the CBB cycle leads to larger plants (Miyagawa et al., 2001; Driever et al., 2017), suggesting the bulk supply of the light reactions are able to meet increased demand from the carbon side. However, if we think about the system dynamically, there are parts of the day that plants are light limited, such as early morning. Under very low light, q_E and PCON are not active, and suppression of them wouldn't likely offer substantial benefit. However, these processes are activated at intermediate light, where there is room to modify feedback regulation to improve photosynthesis (Murchie and Niyogi, 2011). Rather than modify the response of q_E and PCON directly, one approach that has potential is the modification of the metabolic sensor of the ETC, the ATP synthase.

The ATP synthase is the central regulator of photosynthesis. It senses downstream metabolic states and controls D*p* accumulation. For example, when the CBB cycle is substrate limited, the rate constant of proton efflux decreases, leading to the activation of q_E and PCON, even at low light (Kanazawa and Kramer, 2002). The ATP synthase is post-translationally regulated, the most understood of which is the light to dark inactivation of the redox regulated g-subunit (Kramer et al., 1990; Evron et al., 2000; Wu et al., 2007). However, the ATP synthase has also been proposed to be regulated by phosphorylation and protein-protein interactions, but these are not clearly understood (Evron et al., 2000; Bunney et al., 2001; Reiland et al., 2009). Additionally, it has been demonstrated that up to 50% of the ATP synthase content in tobacco leaves is inactive, and the molecular mechanism of this inactivation is completely unknown (Rott et al., 2011). Clearly, more basic research into the regulation of ATP synthase could lead to strategies to delay the onset of feedback regulation leading to q_E and PCON, and increase the integrated efficiency of the light reactions during the course of the day.

While the energetics of target pathways are not the only engineering consideration, we hope we have emphasized the importance of chloroplast energetics in synthetic biology. While some bioengineering strategies are promising, relatively extreme differences from native chloroplast metabolic demands may require secondary engineering to maximize the yield of the desired product. Finally, bioengineering of plants to fine tune energetic responses may lead to improved growth and resilience to stress conditions. Fine tuning should be considered as a strategy for plant improvement in addition to methods that may increase yield more dramatically.

Author contributions

Both authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Funding

This work was supported by the U.S. Department of Energy Office of Science, Basic Energy Sciences under Award DE-FG02-91ER20021.

Acknowledgments

We thank Drs. Nick Fisher and Mauricio Tejera (Michigan State University), and Dr. Charles Haynes (University of British Columbia) for inspiring discussions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Ahmad, N., Michoux, F., Lössl, A. G., and Nixon, P. J. (2016). Challenges and perspectives in commercializing plastid transformation technology. *J. Exp. Bot.* 67, 5945–5960. doi: 10.1093/jxb/erw360

Alishah Aratboni, H., Rafiei, N., Garcia-Granados, R., Alemzadeh, A., and Morones-Ramírez, J. R. (2019). Biomass and lipid induction strategies in microalgae for biofuel production and other applications. *Microb. Cell Fact.* 18, 1–17. doi: 10.1186/s12934-019-1228-4

Alric, J., and Johnson, X. (2017). Alternative electron transport pathways in photosynthesis: a confluence of regulation. *Curr. Opin. Plant Biol.* 37, 78–86. doi: 10.1016/j.pbi.2017.03.014

Armbruster, U., Leonelli, L., Correa Galvis, V., Strand, D., Quinn, E. H., Jonikas, M. C., et al. (2016). Regulation and levels of the thylakoid K+/H+ antiporter KEA3 shape the dynamic response of photosynthesis in fluctuating light. *Plant Cell Physiol.* 57, 1557–1567. doi: 10.1093/pcp/pcw085

Arnon, D. I. (1971). The light reactions of photosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 68, 2883–2892. doi: 10.1016/1011-1344(92)87001-P

Asada, K. (2000). The water-water cycle as alternative photon and electron sinks. *Philos. Trans. R. Soc Lond. B. Biol. Sci.* 355, 1419–1431. doi: 10.1098/rstb.2000.0703

Blankenship, R. (2002). *Molecular mechanisms of photosynthesis*. Ed. R. E. Blankenship (Oxford, UK: Blackwell Science Ltd). doi: 10.1002/9780470758472

Blatti, J. L., Michaud, J., and Burkart, M. D. (2013). Engineering fatty acid biosynthesis in microalgae for sustainable biodiesel. *Curr. Opin. Chem. Biol.* 17, 496–505. doi: 10.1016/j.cbpa.2013.04.007

Bock, R. (2001). Transgenic plastids in basic research and plant biotechnology. *J. Mol. Biol.* 312, 425–438. doi: 10.1006/jmbi.2001.4960

Bock, R. (2015). Engineering plastid genomes: Methods, tools, and applications in basic research and biotechnology. *Annu. Rev. Plant Biol.* 66, 211–241. doi: 10.1146/annurev-arplant-050213-040212

Bock, R. (2022). Transplastomic approaches for metabolic engineering. *Curr. Opin. Plant Biol.* 66, 102185. doi: 10.1016/j.pbi.2022.102185

Boynton, J. E., Gillham, N. W., Harris, E. H., Hosler, J. P., Johnson, A. M., Jones, A. R., et al. (1988). Chloroplast transformation in chlamydomonas with high velocity microprojectiles. *Science* 240, 1534–1538. doi: 10.1126/science.2897716

Bunney, T. D., Van Walraven, H. S., and De Boer, A. H. (2001). 14-3-3 protein is a regulator of the mitochondrial and chloroplast ATP synthase. *Proc. Natl. Acad. Sci. U. S. A.* 9, 4249–4254. doi: 10.1073/pnas.061437498

Casano, L. M., Martín, M., and Sabater, B. (2001). Hydrogen peroxide mediates the induction of chloroplastic ndh complex under photooxidative stress in barley. *Plant Physiol.* 125, 1450–1458. doi: 10.1104/pp.125.3.1450

Chen, J. H., Chen, S. T., He, N. Y., Wang, Q. L., Zhao, Y., Gao, W., et al. (2020). Nuclear-encoded synthesis of the D1 subunit of photosystem II increases photosynthetic efficiency and crop yield. *Nat. Plants* 6, 570–580. doi: 10.1038/s41477-020-0629-z

Cruz, J. A., Sacksteder, C. A., Kanazawa, A., and Kramer, D. M. (2001). Contribution of electric field (Delta psi) to steady-state transthylakoid proton motive force (pmf) *in vitro* and *in vivo*. control of pmf parsing into delta psi and delta pH by ionic strength. *Biochemistry* 40, 1226–1237. doi: 10.1021/bi0018741

De Souza, A. P., Burgess, S. J., Doran, L., Hansen, J., Manukyan, L., Maryn, N., et al. (2022). Soybean photosynthesis and crop yield are improved by accelerating recovery from photoprotection. *Sci. (80-.).* 377, 851–854. doi: 10.1126/science.adc9831

De Wilde, C., Van Houdt, H., De Buck, S., Angenon, G., De Jaeger, G., and Depicker, A. (2000). Plants as bioreactors for protein production: Avoiding the problem of transgene silencing. *Plant Mol. Biol.* 43, 347–359. doi: 10.1023/a:1006464304199

Driever, S. M., Simkin, A. J., Alotaibi, S., Fisk, S. J., Madgwick, P. J., Sparks, C. A., et al. (2017). Increased sbpase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. *Philos. Trans. R. Soc B Biol. Sci.* 372, 20160384. doi: 10.1098/rstb.2016.0384

Edwards, G., and Walker, D. A. (2003). *C3,C4: mechanism, and cellular and environmental regulation, of photosynthesis.* (Oxford: Blackwell Scientific Publications).

Ermakova, M., Lopez-Calcagno, P. E., Raines, C. A., Furbank, R. T., and von Caemmerer, S. (2019). Overexpression of the rieske FeS protein of the cytochrome b 6 f complex increases C4 photosynthesis in setaria viridis. *Commun. Biol.* 2, 314. doi: 10.1038/s42003-019-0561-9

Evron, Y., Johnson, E. A., and Mccarty, R. E. (2000). Regulation of proton flow and ATP synthesis in chloroplasts. *J. Bioenerg. Biomembr.* 32, 501–506. doi: 10.1023/A:1005669008974

Flügge, U. I., Häusler, R. E., Ludewig, F., and Gierth, M. (2011). The role of transporters in supplying energy to plant plastids. *J. Exp. Bot.* 62, 2381–2392. doi: 10.1093/jxb/erq361

Foyer, C. H., Bloom, A. J., Queval, G., and Noctor, G. (2009). Photorespiratory metabolism: Genes, mutants, energetics, and redox signaling. *Annu. Rev. Plant Biol.* 60, 455–484. doi: 10.1146/annurev.arplant.043008.091948

Fu, X., Gregory, L. M., Weise, S. E., and Walker, B. J. (2022a). Integrated flux and pool size analysis in plant central metabolism reveals unique roles of glycine and serine during photorespiration. *Nat. Plants.* 203–222 doi: 10.1038/s41477-022-01294-9

Fu, X., Smith, K., Gregory, L., Roze, L., and Walker, B. (2022b). *Modifying photorespiration to optimize crop performance. Understanding and improving crop photosynthesis*. R. Sharwood. (editor), Understanding and improving crop photosynthesis, Cambridge, UK: Burleigh Dodds Science Publishing, 2023. www.bdspublishing.com.

Garcia-Molina, A., and Leister, D. (2020). Accelerated relaxation of photoprotection impairs biomass accumulation in arabidopsis. *Nat. Plants* 6, 9–12. doi: 10.1038/s41477-019-0572-z

Gilmour, D. J. (2019). Microalgae for biofuel production. *Adv Appl Microbiol.* 109:1–30. doi: 10.1016/bs.aambs.2019.10.001

Hahn, A., Vonck, J., Mills, D. J., Meier, T., and Kühlbrandt, W. (2018). Structure, mechanism, and regulation of the chloroplast ATP synthase. *Science* 360, eaat4318. doi: 10.1126/science.aat4318

HAY, R. K. M. (1995). Harvest index: a review of its use in plant breeding and crop physiology. *Ann. Appl. Biol.* 126, 197–216. doi: 10.1111/j.1744-7348.1995.tb05015.x

Hill, R., and Bendall, F. (1960). Function of the two cytochrome components in chloroplasts: A working hypothesis. *Nature* 186, 136–137. doi: 10.1038/186136a0

Kanazawa, A., and Kramer, D. M. (2002). *In vivo* modulation of nonphotochemical exciton quenching (NPQ) by regulation of the chloroplast ATP synthase. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12789–12794. doi: 10.1073/pnas.182427499

Kaushal, C., Abdin, M. Z., and Kumar, S. (2020). Chloroplast genome transformation of medicinal plant artemisia annua. *Plant Biotechnol. J.* 18, 2155–2157. doi: 10.1111/pbi.13379

Kramer, D. M., Cruz, J. A., and Kanazawa, A. (2003). Balancing the central roles of the thylakoid proton gradient. *Trends Plant Sci.* 8, 27–32. doi: 10.1016/S1360-1385(02)00010-9

Kramer, D. M., and Evans, J. R. (2011). The importance of energy balance in improving photosynthetic productivity. *Plant Physiol.* 155, 70–78. doi: 10.1104/pp.110.166652

Kramer, D. M., Wise, R. R., Frederick, J. R., Alm, D. M., Hesketh, J. D., Ort, D. R., et al. (1990). Regulation of coupling factor in field-grown sunflower: A redox model relating coupling factor activity to the activities of other thioredoxin-dependent chloroplast enzymes. *Photosynth. Res.* 26, 213–222. doi: 10.1007/BF00033134

Kromdijk, J., Głowacka, K., Leonelli, L., Gabilly, S. T., Iwai, M., Niyogi, K. K., et al. (2016). Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Sci. (80-.).* 354, 857–861. doi: 10.1126/science.aai8878

Li, X.-P., Muller-Moule, P., Gilmore, A. M., and Niyogi, K. K. (2002). PsbS-dependent enhancement of feedback de-excitation protects photosystem II from photoinhibition. *Proc. Natl. Acad. Sci. U. S. A.* 99, 15222–15227. doi: 10.1073/pnas.232447699

Lin, M. T., Occhialini, A., Andralojc, P. J., Devonshire, J., Hines, K. M., Parry, M. A. J., et al. (2014). β-carboxysomal proteins assemble into highly organized structures in nicotiana chloroplasts. *Plant J.* 79, 1–12. doi: 10.1111/tpj.12536

Livingston, A. K., Cruz, J. A., Kohzuma, K., Dhingra, A., and Kramer, D. M. (2010). An arabidopsis mutant with high cyclic electron flow around photosystem I (hcef) involving the NADPH dehydrogenase complex. *Plant Cell* 22, 221–233. doi: 10.1105/tpc.109.071084

Long, S. P., Zhu, X.-G., Naidu, S. L., and Ort, D. R. (2006). Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* 29, 315–330. doi: 10.1111/j.1365-3040.2005.01493.x

Ma, F., Jazmin, L. J., Young, J. D., and Allen, D. K. (2014). Isotopically nonstationary 13C flux analysis of changes in arabidopsis thaliana leaf metabolism due to high light acclimation. *Proc. Natl. Acad. Sci. U. S. A.* 111, 16967–16972. doi: 10.1073/pnas.1319485111

Miller, T. E., Beneyton, T., Schwander, T., Diehl, C., Girault, M., McLean, R., et al. (2020). Light-powered CO2 fixation in a chloroplast mimic with natural and synthetic parts. *Sci. (80-.).* 368, 649–654. doi: 10.1126/science.aaz6802

Mitchell, P. (1975). The protonmotive q cycle: a general formulation. *FEBS Lett.* 59, 137–139. doi: 10.1016/0014-5793(75)80359-0

Mitchell, P. (2011). Chemiosmotic coupling in oxidative and photosynthetic phosphorylation. 1966. *Biochim. Biophys. Acta* 1807, 1507–1538. doi: 10.1016/j.bbabio.2011.09.018

Miyagawa, Y., Tamoi, M., and Shigeoka, S. (2001). Overexpression of a cyanobacterial. *Nat. Biotechnol.* 19, 965–969. doi: 10.1038/nbt1001-965

Müller, P., Li, X. P., and Niyogi, K. K. (2001). Non-photochemical quenching. a response to excess light energy. *Plant Physiol.* 125, 1558–1566. doi: 10.1104/pp.125.4.1558

Murchie, E. H., and Niyogi, K. K. (2011). Manipulation of photoprotection to improve plant photosynthesis. *Plant Physiol.* 155, 86–92. doi: 10.1104/pp.110.168831

Nicholls, D., and Ferguson, S. (2002). *Bioenergetics. 3rd ed* (London: Academic Press Inc).

Nimmo, I. C., Barbrook, A. C., Lassadi, I., Chen, J. E., Geisler, K., Smith, A. G., et al. (2019). Genetic transformation of the dinoflagellate chloroplast. *Elife* 8, 1–15. doi: 10.7554/eLife.45292

Niyogi, K. K., Grossman, A. R., and Björkman, O. (1998). Arabidopsis mutants define a central role for the xanthophyll cycle in the regulation of photosynthetic energy conversion. *Plant Cell* 10, 1121–1134. doi: 10.1105/tpc.10.7.1121

Noctor, G., and Foyer, C. H. (1998). A re-evaluation of the ATP:NADPH budget during C3 photosynthesis: A contribution from nitrate assimilation and its associated respiratory activity? *J. Exp. Bot.* 49, 1895–1908. doi: 10.1093/jxb/49.329.1895

Ogren, W. L. (1984). Photorespiration: Pathways, regulation, and modification. *Annu. Rev. Plant Physiol.* 35, 415–442. doi: 10.1146/annurev.pp.35.060184.002215

Ohlrogge, J., and Browse, J. (1995). Lipid biosynthesis. *Plant Cell* 7, 957–970. doi: 10.1105/tpc.7.7.957

O'Neill, E. C., and Kelly, S. (2017). Engineering biosynthesis of high-value compounds in photosynthetic organisms. *Crit. Rev. Biotechnol.* 37, 779–802. doi: 10.1080/07388551.2016.1237467

Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., Blankenship, R. E., Bock, R., et al. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci. U. S. A.* 112, 8529–8536. doi: 10.1073/pnas.1424031112

Paul, M. J. (2021). What are the regulatory targets for intervention in assimilate partitioning to improve crop yield and resilience? *J. Plant Physiol.* 266, 153537. doi: 10.1016/j.jplph.2021.153537

Peterhansel, C., Horst, I., Niessen, M., Blume, C., Kebeish, R., Kürkcüoglu, S., et al. (2010). Photorespiration. *Arab. B.* 8, e0130. doi: 10.1199/tab.0130

Reiland, S., Messerli, G., Baerenfaller, K., Gerrits, B., Endler, A., Grossmann, J., et al. (2009). Large-Scale arabidopsis phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. *Plant Physiol.* 150, 889–903. doi: 10.1104/pp.109.138677

Rott, M., Martins, N. F., Thiele, W., Lein, W., Bock, R., Kramer, D. M., et al. (2011). ATP synthase repression in tobacco restricts photosynthetic electron transport, CO2 assimilation, and plant growth by overacidification of the thylakoid lumen. *Plant Cell* 23, 304–321. doi: 10.1105/tpc.110.079111

Ruf, S., Kroop, X., and Bock, R. (2021). Chloroplast transformation in arabidopsis. *Curr. Protoc.* 1, 1–24. doi: 10.1002/cpz1.103

Schwander, T., Borzyskowski, L. S., Von, Burgener, S., Socorro Cortina, N., and Erb, T. J. (2016). A synthetic pathway for the fixation of carbon dioxide *in vitro*. *Sci. (80-.).* 354, 900–904. doi: 10.1126/science.aah5237.A

Seelert, H., Dencher, N. A., and Müller, D. J. (2003). Fourteen protomers compose the oligomer III of the proton-rotor in spinach chloroplast ATP synthase. *J. Mol. Biol.* 333, 337–344. doi: 10.1016/j.jmb.2003.08.046

Sharkey, T. D., and Weise, S. E. (2016). The glucose 6-phosphate shunt around the Calvin-Benson cycle. *J. Exp. Bot.* 67, 4067–4077. doi: 10.1093/jxb/erv484

South, P. F., Cavanagh, A. P., Liu, H. W., and Ort, D. R. (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Sci. (80-.).* 363, 0–10. doi: 10.1126/science.aat9077

Strand, D. D., Fisher, N., and Kramer, D. M. (2016). "Distinct energetics and regulatory functions of the two major cyclic electron flow pathways in chloroplasts," in *Chloroplasts: Current research and future trends*. Ed. H. Kirchhoff (United Kingdom: Caister Academic Press), 89–100. doi: 10.21775/9781910190470.04

Strand, D. D., Fisher, N., and Kramer, D. M. (2017). The higher plant plastid NAD(P)H dehydrogenase-like complex (NDH) is a high efficiency proton pump that increases ATP production by cyclic electron flow. *J. Biol. Chem.* 292, 11850–11860. doi: 10.1074/jbc.M116.770792

Strand, D. D., Livingston, A. K., Satoh-Cruz, M., Froehlich, J. E., Maurino, V. G., and Kramer, D. M. (2015). Activation of cyclic electron flow by hydrogen peroxide *in vivo*. *Proc. Natl. Acad. Sci. U. S. A.* 112, 5539–5544. doi: 10.1073/pnas.1418223112

Suorsa, M., Järvi, S., Grieco, M., Nurmi, M., Pietrzykowska, M., Rantala, M., et al. (2012). PROTON GRADIENT REGULATION5 is essential for proper acclimation of arabidopsis photosystem I to naturally and artificially fluctuating light conditions. *Plant Cell* 24, 2934–2948. doi: 10.1105/tpc.112.097162

Svab, Z., and Maliga, P. (1993). High-frequency plastid transformation in tobacco by selection for a chimeric aadA gene. *Proc. Natl. Acad. Sci. U. S. A.* 90, 913–917. doi: 10.1073/pnas.90.3.913

Verma, S., and Kuila, A. (2020). Involvement of green technology in microalgal biodiesel production. *Rev. Environ. Health* 35, 173–188. doi: 10.1515/reveh-2019-0061

Walker, B. J., Kramer, D. M., Fisher, N., and Fu, X. (2020). Flexibility in the energy balancing network of changing environmental conditions. *Plants* 9, 301. doi: 10.3390/plants9030301

Wang, L. M., Shen, B. R., Li, B., Zhang, C. L., Lin, M., Tong, P. P., et al. (2020). A synthetic photorespiratory shortcut enhances photosynthesis to boost biomass and grain yield in rice. *Mol. Plant* 13, 1802–1815. doi: 10.1016/j.molp.2020.10.007

Wieloch, T., and Sharkey, T. D. (2022). Compartment-specific energy requirements of photosynthetic carbon metabolism in camelina sativa leaves. *Planta* 255, 1–10. doi: 10.1007/s00425-022-03884-5

Wu, G., Ortiz-Flores, G., Ortiz-Lopez, A., and Ort, D. R. (2007). A point mutation in atpC1 raises the redox potential of the arabidopsis chloroplast ATP synthase gamma-subunit regulatory disulfide above the range of thioredoxin modulation. *J. Biol. Chem.* 282, 36782–36789. doi: 10.1074/jbc.M707007200

Yu, S. M., Lo, S. F., and Ho, T. H. D. (2015). Source-sink communication: Regulated by hormone, nutrient, and stress cross-signaling. *Trends Plant Sci.* 20, 844–857. doi: 10.1016/j.tplants.2015.10.009

Zhao, L., Chang, W., Xiao, Y., Liu, H., and Liu, P. (2013). Methylerythritol phosphate pathway of isoprenoid biosynthesis. *Annu. Rev. Biochem.* 82, 497–530. doi: 10.1146/annurev-biochem-052010-100934

frontiers | Frontiers in Plant Science

TYPE Original Research
PUBLISHED 03 July 2023
DOI 10.3389/fpls.2023.1133299

OPEN ACCESS

EDITED BY
Zhi-Yan (Rock) Du,
University of Hawaii at Manoa,
United States

REVIEWED BY
Pan Liao,
Hong Kong Baptist University,
Hong Kong SAR, China
Mo-Xian Chen,
Guizhou University, China

*CORRESPONDENCE
Rui Alves
rui.alves@udl.cat

RECEIVED 28 December 2022
ACCEPTED 30 May 2023
PUBLISHED 03 July 2023

CITATION
Basallo O, Perez L, Lucido A, Sorribas A, Marin-Saguino A, Vilaprinyo E, Perez-Fons L, Albacete A, Martínez-Andújar C, Fraser PD, Christou P, Capell T and Alves R (2023) Changing biosynthesis of terpenoid percursors in rice through synthetic biology. *Front. Plant Sci.* 14:1133299. doi: 10.3389/fpls.2023.1133299

Changing biosynthesis of terpenoid percursors in rice through synthetic biology

Orio Basallo[1,2], Lucia Perez[3,4], Abel Lucido[1,2], Albert Sorribas[1,2], Alberto Marin-Saguino[1,2], Ester Vilaprinyo[1,2], Laura Perez-Fons[5], Alfonso Albacete[6,7], Cristina Martínez-Andújar[6], Paul D. Fraser[5], Paul Christou[3,4,8], Teresa Capell[3,4] and Rui Alves[1,2]*

[1]Systems Biology Group, Department Ciències Mèdiques Bàsiques, Faculty of Medicine, Universitat de Lleida, Lleida, Spain, [2]Institut de Recerca Biomedica de Lleida (IRBLleida), Lleida, Spain, [3]Applied Plant Biotechnology Group, Department de Producció Vegetal I Ciència Florestal, Escola Tècnica Superior d'Enginyeria Agroalimentària i Forestal i de Veterinària (ETSEAFiV), Universitat de Lleida, Lleida, Spain, [4]Agrotecnio Centres de Recerca de Catalunya (CERCA) Center, Lleida, Spain, [5]School of Biological Sciences, Royal Holloway University of London, Egham Hill, United Kingdom, [6]Departament of Plant Nutrition, Center of Edaphology and Applied Biology of the Segura (CEBAS), Consejo Superior de Investigaciones Científicas (CSIC), Universidad de Murcia, Murcia, Spain, [7]Department of Plant Production and Agrotechnology, Institute for Agri-Food Research and Development of Murcia, Murcia, Spain, [8]ICREA, Catalan Institute for Research and Advanced Studies, Barcelona, Spain

Many highly valued chemicals in the pharmaceutical, biotechnological, cosmetic, and biomedical industries belong to the terpenoid family. Biosynthesis of these chemicals relies on polymerization of Isopentenyl di-phosphate (IPP) and/or dimethylallyl diphosphate (DMAPP) monomers, which plants synthesize using two alternative pathways: a cytosolic mevalonic acid (MVA) pathway and a plastidic methyleritritol-4-phosphate (MEP) pathway. As such, developing plants for use as a platform to use IPP/DMAPP and produce high value terpenoids is an important biotechnological goal. Still, IPP/DMAPP are the precursors to many plant developmental hormones. This creates severe challenges in redirecting IPP/DMAPP towards production of non-cognate plant metabolites. A potential solution to this problem is increasing the IPP/DMAPP production flux *in planta*. Here, we aimed at discovering, understanding, and predicting the effects of increasing IPP/DMAPP production in plants through modelling. We used synthetic biology to create rice lines containing an additional ectopic MVA biosynthetic pathway for producing IPP/DMAPP. The rice lines express three alternative versions of the additional MVA pathway in the plastid, in addition to the normal endogenous pathways. We collected data for changes in macroscopic and molecular phenotypes, gene expression, isoprenoid content, and hormone abundance in those lines. To integrate the molecular and macroscopic data and develop a more in depth understanding of the effects of engineering the exogenous pathway in the mutant rice lines, we developed and analyzed data-centric, line-specific, multilevel mathematical models. These models connect the effects of variations in hormones and gene expression to changes in macroscopic plant phenotype and metabolite concentrations within the MVA and MEP pathways of WT and mutant rice lines. Our models allow us to predict how an exogenous IPP/DMAPP biosynthetic pathway affects the flux of terpenoid precursors. We also quantify the long-term effect of plant hormones on the dynamic behavior of IPP/DMAPP biosynthetic pathways in seeds, and

predict plant characteristics, such as plant height, leaf size, and chlorophyll content from molecular data. In addition, our models are a tool that can be used in the future to help in prioritizing re-engineering strategies for the exogenous pathway in order to achieve specific metabolic goals.

KEYWORDS

metabolic engineering, mathematical modelling, multi level modelling, MVA (mevalonic acid) pathway, MEP pathway, terpenoid synthetic biology

1 Introduction

Terpenoids are a family of molecules with more than 22,000 different natural products (Harborne et al., 1991; Tetali, 2019; Zhou and Pichersky, 2020; Navale et al., 2021). Some family members have various crucial biological functions. For example, in plants, they work as hormones (gibberellin, abscisic acid, etc.), photosynthetic pigments (chlorophyll, phytol, carotenoids), electron carriers (ubiquinone, plastoquinone), mediators of the assembly of polysaccharides (polyprenyl phosphates) and structural components of membranes (phytosterols). They are also used for other purposes, such as antibiotics, herbivore repellents, toxins and pollinator attractants (Mcgarvey and Croteau, 1995).

Plants synthesize terpenoids from two metabolic precursors: Isopentenyl di-phosphate (IPP) and dimethylallyl diphosphate (DMAPP). Two compartmentally separated pathways synthesize these precursors (Figure 1). The mevalonic acid (MVA) pathway converts acetyl-CoA (Ac-CoA) to IPP and DMAPP. This pathway is mostly cytosolic, with a couple of reactions taking place in the peroxisome. The MVA pathway starts with the condensation of acetyl-CoA, a product of glycolysis, catalyzed by acetoacetyl-CoA thiolase and HMG-CoA synthase, followed by the conversion of HMG-CoA to mevalonate through HMG-CoA reductase. Mevalonate is subsequently phosphorylated and decarboxylated to yield IPP, which can be isomerized to DMAPP by the action of isopentenyl diphosphate isomerase (IDI). IPP and DMAPP are then used in the synthesis of phytosterols and ubiquinone (Mcgarvey and Croteau, 1995). The enzyme 3-hydroxy-3-methylglutaril-CoA reductase (HMGR) is a key enzyme in the regulation of the MVA pathway (Schaller et al., 1995). The second terpenoid-producing pathway is known as the methyleritritol-4-phosphate (MEP) pathway. This pathway is compartmentalized in plastids. In this pathway, glyceraldehyde 3-phosphate (G3P) and pyruvate derived from the Calvin cycle serve as the primary carbon sources for IPP and DMAPP production. The MEP pathway involves a series of

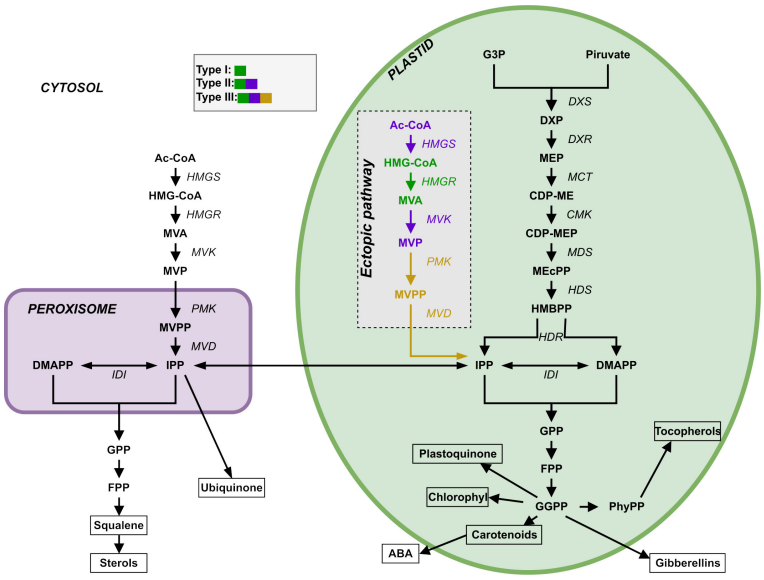


FIGURE 1
Representation of the two terpenoid biosynthesis pathways plus the ectopic pathway, the MVA pathway (left, cytosol and peroxisome) and the MEP pathway (right, plastid). DXR: DXP reductoisomerase; MCT: 2-C-methyl-D-erythrtle 4-phosphate cytidylyl transferase; CDP-ME: 4-(Citidine 5'-difosfo)-2-C-methyl-D-eritritol; CMK: 4-difosfocitidil-2-C-methyl-D-erythrtol kinase; CDP-MEP: 2-Fosfo-4-(cytidine 5'- diphospho)-2-C-methyl-D-eritritol; MDS: 2-C-methyl-D-eritritol 2,4-cyclodifosphate synthase; MEcPP: 2-C-methyl-D-eritritol 2,4-cycdiphosphate; HDS: 4-hydroxy-3-methylbut-2-en-1-il diphosphate synthase; HMBPP: 4-hydroxy-3-methylbut-2-in-1-il diphosphate; HDR: 4-hydroxy-3-methylbut-2-en-1-il diphosphate reductase; IDI: isopentenyl diphosphate Delta-isomerase; PhyPP: phytyl diphosphate.

enzymatic steps catalyzed by various enzymes, including 1-deoxy-D-xylulose 5-phosphate synthase (DXS), 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR), and other downstream enzymes. It is responsible for the production of carotenoids, lateral chains of chlorophylls, plastoquinone, abscisic acid (ABA) and tocopherols (vitamin E, precursors and derivatives) (Eisenreich et al., 2001). Appendix S1 presents a more detailed description of both pathways.

While both pathways function independently, there is ample evidence of crosstalk between them (Hemmerlin et al., 2003; Hemmerlin, 2013). There is evidence for the exchange of some metabolic intermediates of the two pathways between compartments (Bick and Lange, 2003; Hemmerlin et al., 2003; Laule et al., 2003; Hemmerlin et al., 2006). The first intermediate of the MEP pathway, DXP, can diffuse between the plastid and the cytoplasm (Hemmerlin et al., 2003; Page et al., 2004; Lange et al., 2015). At the level of IPP and DMAPP, this exchange was measured to occur mainly in the plastid-to-cytoplasm direction, promoted by a one-way symport system (Bick and Lange, 2003; Dudareva et al., 2005). The direction of this metabolic exchange between cellular compartments may depend on physiological state and species. There is lack of convincing evidence that other intermediates of both pathways can diffuse between the two compartments (Hemmerlin, 2013). Hemmerlin et al. (2012) made an extensive review of the literature covering the metabolic pathways themselves (networks, regulation, biological advantage of having two separated pathways, etc.), crosstalk between pathways and the potential of terpenoid biosynthesis in bioengineered plants in biotechnology.

In recent years, synthetic biology has emerged as a powerful tool for engineering production of isoprenoids, mostly into *Saccharomyces cerevisiae* and other microbial hosts (Ro et al., 2006; Ajikumar et al., 2010; Jiang et al., 2017; Kang et al., 2017; Cravens et al., 2019; Luo et al., 2019; Gülck et al., 2020; Srinivasan and Smolke, 2020; Liew et al., 2022; Zhang et al., 2022). While microbial production is very attractive due to the speed and easiness of genetically manipulating microorganisms, plant production should remain center stage if we are to progress to a circular bio-economy (Shih, 2018). Some of the earliest examples include the modifications made to maize (Naqvi et al., 2009) and rice (Ye et al., 2000), in order to increase their vitamin content. In general, modulating metabolite levels in plants is attempted through varying combinations of the following strategies: increasing enzymatic activity, increasing availability of upstream precursors, blocking leakage of the compound by gene silencing, or inducing metabolite storage in a compartment (Zorrilla-López et al., 2013; Kotopka et al., 2018; Maeda, 2019; Zhu et al., 2021; Yang et al., 2022). Examples of plant modification to enhance production of many complex isoprenoids became increasingly common (reviewed in Liao et al. (2016)). For example, Kumar et al. modified the chloroplast genome of tobacco leaves in a non-transmissible way to code an exogenous MVA pathway (Kumar et al., 2012). A more recent example is the enhancement in production of sesquiterpene precursor FPP of the MEP pathway in tomato fruit (Chen et al., 2023). As such, plants are a coveted target for engineering pathways that produce high value terpenoid chemicals (Fuentes et al., 2016; Georgiev et al., 2018; Gülck et al., 2020; Grzech et al., 2023).

Pérez et al. (2022) used synthetic biology to create rice plants that have a stable and transmissible ectopic plastidial MVA pathway that coexists with the native MVA and MEP pathways and is expressed in endosperm. The goal of that study was to circumvent the regulation of the native MVA pathway and test the possibility of producing an excess of terpenoids precursors that could feed the biosynthesis of highly valuable terpenoids. They introduced three different combinations of exogenous WR1, HMGS, HMGR, MVK, PMK and MVD (Figure 1) genes encoding plastid-targeted enzymes, collecting transcriptomic, metabolic, and phenotypic data for the resulting mutant lines. The WR1 gene is a transcription factor that induces the expression of genes related to plastid glycolysis and fatty acid biosynthesis. HMGS codes for hydroxymethylglutaryl-CoA synthase, HMGR codes for a 3-hydroxy-3-methylglutaryl-coenzyme A reductase, MVK codes for a mevalonate kinase, PMK codes for a phosphomevalonate kinase, and MVD codes for a diphosphomevalonate decarboxylase.

Here, we further study the effect that adding this pathway has on rice by creating and characterizing new rice lines with alternative versions of the exogenous MVA pathway. Subsequently, we use multilevel mathematical modeling to integrate the data of all mutant lines, and predict the effect of genome modification on the concentrations of metabolic intermediates and on the fluxes going through the MEP, MVA, and ectopic plastidial MVA pathways. We also quantify the long-term effect of plant hormones on the dynamic behavior of IPP/DMAPP biosynthetic pathways in seeds, and predict plant characteristics, such as plant height, leaf size, and chlorophyll content from molecular data. In addition, our models are a tool that can be used in the future to help in prioritizing re-engineering strategies for the exogenous pathway in order to achieve specific metabolic goals.

2 Materials and methods

2.1 WT and mutant rice lines

We created three types of mutant rice lines using the procedures described in Pérez et al. (2022). Mutant Type I had exogenous HMGR; Mutant Type II had exogenous HMGS, HMGR and MVK; and Mutant Type III had exogenous HMGS, HMGR, MVK, PMK and MVD. We placed these six transgenes (*BjHMGS*, *tHMGR*, *CrMK*, *CrPMK*, and *CrMVD*) in three independent expression cassettes driven by endosperm-specific promoters. A transit peptide inserted at the beginning of the five enzymes in the MVA pathway directs them to the plastid. See Pérez et al. (2022) and Supplementary Appendix S1 for the full details.

To create the mutant lines, we bombarded seven-day-old mature zygotic rice embryos (*Oryza sativa* cv. EYI105) with gold particles coated with the transformation vectors. We recovered transgenic plantlets and regenerated and hardened them off in soil. Genomic DNA was isolated from the callus and leaves of regenerated plants to confirm presence of the *BjHMGS*, *tHMGR*, *CrMK*, *CrPMK*, *CrMVD* and *OsWR1* through PCR (Pérez et al., 2022). We recovered 12 independent type I mutant lines, 10 independent type II mutant lines, and 12 independent mutant lines.

2.2 Hormone determination, gene expression, and plant phenotypes

We analyzed all rice lines after 12 weeks of growing in soil.

For each line we measured chlorophyll levels and analyzed the cytokinins trans-zeatin, zeatin riboside and isopentenyl adenine (iP); the gibberellins GA1, GA3 and GA4; the auxin indole-3-acetic acid (IAA); ABA; salicylic acid; jasmonic acid; and the ethylene precursor 1-aminocyclopropane-1-carboxylic acid (ACC) as described in Pérez et al. (2022) and Supplementary Appendix S1.

We also measured gene expression for *BjHMGS*, *tHMGR*, *CrMK*, *CrPMK*, *CrMVD*, and the endogenous MVA and MEP pathway genes *OsHMGS*, *OsHMGR*, *OsMK*, *OsPMK*, *OsMVD*, *OsDXS*, *OsDXR*, *OsMCT*, *OsCMK*, *OsMDS*, *OsHDS*, *OsHDR* and *OsIPPI* using qRT-PCR as described in Appendix S1.

We counted the number of leaves, measured the height of the plants from the base of the stem to the maximum extension of the flag leaf, and measured the length and maximum width of the last expanded leaf as described in Pérez et al. (2022). We used a SPAD meter at six points on the last expanded leaf to quantify leaf chlorophyll. We multiplied length and maximum width of the last expanded leaf by a correction factor of 0.75 to estimate leaf area.

All experimental measurements are provided in Supplementary Data S1 and described in the results section of Appendix S1.

2.3 Mathematical modeling formalism

We used ordinary differential equation systems to model the biosynthesis of IPP/DMAPP. The mathematical formalism used to describe the flux dynamics is the saturating formalism (Sorribas et al., 2007; Alves et al., 2008). This formalism allows us to approximate the kinetics of any given reaction using a rational expression, where parameters have physical interpretations that are analogous to those found in classical enzyme kinetics rate expressions. In this formalism, we approximate the rate of a reaction in an inverse space at an operating point by:

$$v \approx \frac{V \prod_{i=1}^{m} x_i}{\prod_{i=1}^{m} (K_i + x_i) + \prod_{b=1}^{p} (x_b + K_b)} \quad (1)$$

V parameters represent apparent saturation rate constants for the reactions. Ki parameters represent apparent binding constants for the substrate(s) or inhibitor(s) of the reaction. While no activators were considered in our model, these can also be included using this formalism.

2.4 The endogenous MVA and MEP pathways

We modelled the wild type IPP and DMAPP production (i.e., the endogenous MVA and MEP pathways), using the canonical reaction set for each pathway, shown in Figure 1. We modelled the kinetics of each process, as well as those for the exchange fluxes of IPP and DMAPP between cytoplasm and plastid, using the rate expressions in Supplementary Tables S1–S3. We assume that the organism is able to maintain homeostasis of Acetyl-CoA and Acetoacetyl-CoA.

2.5 The ectopic MVA pathway in plastid

To model type I mutants, we added the reaction that transforms $HMG\text{-}CoA_{pl}$ into MVA_{pl} to the plastid, as well as the cytoplasm-plastid exchange reactions for these two metabolites (Supplementary Tables S3–S4).

We extended the model for type I mutants to create the model for type II mutants. We included the reactions catalyzed by HMGS and MVK to the plastid compartment (Supplementary Table S4). Corresponding compartment exchange reactions for the substrates and products of these enzymes are also added (Supplementary Table S3).

We extended the model for type II mutants to create the model for type III mutants. We added the metabolite MVPP to the plastid by including the reaction that synthesize it. We also added the reaction that transforms $MVPP_{pl}$ into IPP_{pl}, (Supplementary Table S4). The plastid-cytoplasm exchange flux of $MVPP_{pl}$ is included as well (Supplementary Table S3).

2.6 Exchange of MVA and MEP pathway metabolites between the cytoplasm and the plastid

Under physiological conditions, IPP and DMAPP mostly flow from the plastid into the cytosol (Bick and Lange, 2003). We implemented this observation by assuming that metabolites flow from the plastid to the cytosol at ten times the rate of the import reaction from the cytosol. Bick and Lange (Bick and Lange, 2003) also reported that other pathway intermediates were not actively transported between the two compartments. Other studies confirm this observation (for example, Wright et al. (2014)). However, those same studies show that in mutants overexpressing DXS, there is a second pool of MEcPP outside the chloroplast. In addition early intermediates of the MVA pathway can be found in the plastid space (Schneider et al., 1977).

Introducing an MVA pathway into the plastid as we did, may cause changes in the flux of MVA intermediates between the plastid and the cytoplasm. As such, we allowed for the possibility that HMG-CoA, MVA, MVP and MVPP enter and leave the plastid, albeit at very slow rates. Supplementary Table S3 summarizes all reactions of material interchanged between plastid and cytosol.

2.7 Assembling the ordinary differential equation models for each type of rice

Each metabolite has its own differential equation in the model. The kinetic rate function, f_j, for each process that produces a metabolite M appears as a positive term in the differential equation that determines the dynamic behavior of that metabolite. Similarly, the kinetic rate function, f_k, for each process

that consumes a metabolite appears as a negative term in the differential equation that determines the dynamic behavior of that metabolite:

$$\frac{dM_i}{dt} = \sum_j f_j - \sum_k f_k \quad (2)$$

For the wild type and for each mutant type we assemble a type-specific system of ordinary differential equations (ODEs) that describes the dynamic behavior of all metabolites in the system. As such, we have four different ODE model types. These four models describe the dynamic behavior of all metabolites in the system in a type-specific manner.

2.8 Estimating rate constants, metabolite concentrations and variations in enzyme activity for each line

Supplementary Tables S5 presents the basal kinetic constants for each reaction in the four model types. Supplementary Table S6 presents the concentrations for the independent variables of the four model types.

To personalize the models and make them line specific we need to weigh the rate constants of the relevant type-specific model by the variations in enzyme activity of the individual line of interest. To do so we searched the literature for information about the correlation between changes in gene expression and enzyme activities in the MEP and MVA pathways. As we found no such information, we modeled variations in the enzyme activities of the mutant lines as described in Comas et al. (2016): changes in gene expression with respect to the WT are assumed to be proportional to changes in protein activity. This is the simplest possible assumption about the relationship between changes in gene expression and changes in enzyme activity.

We implement this assumption in the models by explicitly considering the enzymes that catalyze each reaction in the rate expressions. As $V_{max} \approx k_{cat}Enzyme$, the model for the WT sets the enzyme activity to be 1 (the basal state). As we model mutant lines, we assume that changes in gene expression are proportional to changes in enzyme activity. Thus,

$$Enzyme_{line\ i} = Enzyme_{wild\ type} \times (Gene\ \ expression_{line\ i}/Gene\ \ expression_{wild\ type}) \quad (3)$$

In the end we obtained one line-specific model for the wild type rice, 12 line specific models for type I mutant lines, 10 line specific models for type II mutants, and 12 line specific models for type III mutants.

2.9 Stability analysis

While many biological phenomena are rhythmic, overall, biological systems survive because they can achieve homeostasis (Wang et al., 2022). In other words, metabolism remains buffered and stable against normal environmental fluctuations. This occurs for the MEP pathway (Wright et al., 2014). Mathematically this situation is described by a stable steady state. It is well known that, when modeling homeostatic situations, lack of stability is a good diagnostic tool for model incompleteness (Savageau, 1975; Grimbs et al., 2007; Schmidt et al., 2011; Voit, 2013). For example, Ni and Savageau (1996a; 1996b) used this type of diagnostic tool to predict regulatory interactions that could stabilize a model of the red blood cell metabolism. Because of this, we perform stability analysis of the line-specific rice models in order to identify possible model improvements that can stabilize unstable steady states.

An efficient way to assess stability is by calculating the eigenvalues of the Jacobian matrix of the ODE system, which are complex numbers (Voit, 2013). If the real parts of all eigenvalues are negative, the system is stable. Otherwise, the system is unstable. The Jacobian matrix is constructed by taking the partially derivatives of the right-hand side of the ODEs (f_i) with respect to each state variable (x_j), as shown in Eq. 4.

$$J = D_x f = f_x = \frac{\partial f_i}{\partial x_j} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{pmatrix} \quad (4)$$

2.10 Sensitivity analysis

In addition to steady state stability, another tool for model diagnostic is steady state robustness. Reasonable models generate steady states that are robust, and have low sensitivity to parameter changes (Savageau, 1975; Voit, 2013). Sensitivity measures how much a dependent variable or output changes when a parameter is altered (Comas et al., 2016). Parameters with high sensitivities tend to identify where information may be incomplete or inaccurate.

As such, we performed a sensitivity analysis to identify which steps of the pathway could have additional regulation that we were ignoring. We calculated logarithmic, or relative, steady-state parameter sensitivities, which measure the "relative change in a system variable (X) that is caused by a relative change in a parameter (p)" (Voit, 1991):

$$\bar{S}(X,p) = \frac{\partial X/X}{\partial p/p} = \frac{\partial \log X}{\partial \log p} \quad (5)$$

This sensitivity analysis generates a matrix of sensitivities for each line. Each element of the matrix (of dimensions $m \times n$) represents the sensitivity $S_{i,j}$ of metabolite M_i to parameter p_j. To facilitate visualization of sensitivity analysis results and comparison between lines, we compressed the sensitivity analysis matrices for each line in two ways.

First, to see how much a line is sensitive to a certain parameter over all metabolites (variables of the system) we calculate the following index (Comas et al., 2016):

$$S'_j = \frac{\sqrt{\sum_{i=1}^{n} S_{i,j}^2}}{\sqrt{n}} \quad (6)$$

In other words, we calculate the size (or Euclidean norm) of the vector whose components are the sensitivity of each metabolite to parameter p_j, normalized by the number of metabolites in each mutant line (for example, models for Type I lines have 16 metabolites, while those for Type III have 18). We use the Euclidean norm of the sensitivity vectors as a way to represent aggregate sensitivities with a single metric to facilitate visual representation and analysis. Further, and because the number of metabolites increases from type I to type II and from type II to type III models, we make S'_j comparable between models by normalizing it by the number of metabolites considered in the model. Second, to see how sensitive a metabolite is to all parameters in a line we calculate the following index:

$$S'_i = \frac{\sqrt{\sum_{j=1}^{m} S_{i,j}^2}}{\sqrt{m}} \quad (7)$$

As with Eq 6, we calculate the size (or Euclidean norm) of the vector whose components are the sensitivity of the same metabolite M_i to each parameter p_j, normalized by the number of parameters in each mutant line. We use the Euclidean norm of the sensitivity vectors as a way to represent aggregate sensitivities with a single metric to facilitate visual representation and analysis. Further, and because the number of parameters increases from type I to type II and from type II to type III models, we make S'_j comparable between models by normalizing it by the number of parameters considered in the model.

2.11 Investigating hormone influence

To investigate if we could use plant hormone levels as predictors of dynamic behavior in IPP/DMAPP biosynthesis in seeds we performed correlation analysis between hormones and metabolites, as well as genes, as described in Section 1.9 of Appendix S1. Significant effects were then included in the ODE models using one of two possible formalisms:

$$M_i = \alpha H_j^{g_{ij}}, \quad (8)$$

$$M_i = \alpha \left(\frac{H_j}{K + H_j}\right)^{g_{ij}} \quad (9)$$

where $M_i$is the concentration of metabolite i in the model, H_j is the level of hormone j at twelve weeks, and α and g are constants. We chose between the two alternatives in the following way. First we adjust a linear model of Log[M_i] as a function of Log[H_j]. A combination of low adjusted R^2 and high $|g_{ij}|$ suggests a potentially strong influence of the hormone levels on metabolite concentrations (high d $|g_{ij}|$) over a small range of hormone levels (low adjusted R^2). In this situation, we assumed a saturation effect and used Eq 9 to model hormone influence on metabolite production and consumption. Otherwise, we used Eq 8, as it uses a smaller number of parameters and minimizes the possibility of overfitting the model to the data. The threshold for selecting the one or the other formalism was set at 0.5 for the ratio $|R^2_{adj}/g|$ .If $|R^2_{adj}/g| > 0.5$ we use the power law formalism. The right-hand side of the equations modify the ODEs by multiplying the terms that involve production/consumption of the metabolite and involving those enzymes whose genes levels also correlate to hormone levels, in a way that makes the observed correlations affect the production and consumption rates of the metabolites.

For a more detailed procedure, see the Supplementary Appendix S1.

We note that, when hormone levels were below the experimental detection threshold, we reverted the kinetic expression presented in Supplementary Table S7 to the original model, using a piece-wise approximation to solve the differential equations.

2.12 Phenotype models

We used a form of forward stepwise regression (Efroymson, 1960) to investigate how the different phenotypic variables might be predicted from hormones levels, gene expression and metabolite concentrations. We analyzed the following plant phenotypic characteristics: *Height*, number of *Leaves*, *Leaf Length*, *Leaf Width*, and *Chlorophyl* levels. We split experimental data according to mutant type, so that the analysis and model building was performed three times, one for each mutant type. We investigate phenotype as a function of the predictor variables gene expression, hormone levels, and metabolite concentrations.

The first step of the regression analysis was building independent linear models with one predictor variable.

The second step of the regression analysis was to select the predictor variables that had a significant ($\alpha = 0.05$) effect on the phenotype and whose model had an adjusted R^2 greater than 0.2. If only one model has a significant effect, we would choose that one. If more than one predictor variable has a significant effect in explaining the predicted variable, we chose the model for the variable with the highest adjusted R^2. If the adjusted R^2 is similar between models, we chose the model with the lowest AICc (AIC corrected for small sample sizes) score. The lower this score, the lower the chance that a model over fits the observations. If the AICc is similar for more than one model, we selected the predictor variable with the highest adjusted R^2. At this stage, we have a one variable model.

The third step of the analysis was to create models with all possible combinations of predictor variables where one of the elements of the pair is the predictor variable selected in step one. We then selected the best two variable models as described in the previous paragraph, while making sure they are not collinear.

We repeated steps two and three and stopped when adding a new variable did not improve the explanatory power of the model. Thus, for a given set of significant predictors $\{x_1, x_2, \ldots, x_n\}$, the model would be:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n \quad (10)$$

We used forward stepwise regression instead of the more traditional multilinear modelling approach that starts from Eq. 10 and eliminates all variables that have no effect because the number of data points is smaller than the number of parameters to fit to the full multilinear model.

3 Results

3.1 Mathematical description of IPP and DMAPP biosynthetic pathways

The full mathematical description of IPP/DMAPP biosynthesis consists of 14, 16, 17 and 18 differential equations for the wild type (WT), Mutant Type I (MT-I), Mutant Type II (MT-II) and Mutant Type III (MT-III), respectively. Eq 11 shows the overall ODE systems for the four model types:

$$\begin{aligned}
\dot{HMG\text{-}CoA}_{cyt} &= r_1 + r_{33} - r_2 - r_{32} \\
\dot{HMG\text{-}CoA}_{pl} &= r_{27} + r_{32} - r_{28} - r_{33} \\
\dot{MVA}_{cyt} &= r_2 + r_{35} - r_3 - r_{34} \\
\dot{MVA}_{pl} &= r_{28} + r_{34} - r_{29} - r_{35} \\
\dot{MVP}_{cyt} &= r_3 + r_{37} - r_4 - r_{36} \\
\dot{MVP}_{pl} &= r_{29} + r_{36} - r_{30} - r_{37} \\
\dot{MVPP}_{cyt} &= r_4 + r_{39} - r_5 - r_{38} \\
\dot{MVPP}_{pl} &= r_{30} + r_{38} - r_{31} - r_{39} \\
\dot{IPP}_{cyt} &= r_5 + r_7 + r_9 - r_6 - r_8 - 4r_{20} \\
\dot{IPP}_{pl} &= r_8 + r_{16} + r_{19} + r_{31} - r_9 - r_{18} - 3r_{22} \\
\dot{DMAPP}_{cyt} &= r_6 + r_{26} - r_7 - 2r_{20} - r_{25} \\
\dot{DMAPP}_{pl} &= r_{17} + r_{18} + r_{25} - r_{19} - r_{22} - r_{26} \\
\dot{DXP} &= r_{10} - r_{11} \\
\dot{MEP} &= r_{11} - r_{12} \\
\dot{CDP\text{-}ME} &= r_{12} - r_{13} \\
\dot{CDP\text{-}MEP} &= r_{13} - r_{14} \\
\dot{MEcPP} &= r_{14} - r_{15} \\
\dot{HMBPP} &= r_{15} - r_{16} - r_{17}
\end{aligned} \quad (11)$$

WT: Black

MT-I: Black + green

MT-II: Black + green + purple

MT-III: Black + green + purple + yellow

3.2 MVA and MEP homeostasis is robust in the WT line

Table 1 provides the concentrations of each metabolite estimated from the model in the WT line. The system can achieve homeostasis (stable steady state in mathematical nomenclature). Stable steady states have negative real parts for the eigenvalues of the system's Jacobean matrix (Table 2). The model also estimates that pathway substrates (HMG-CoA and DXP) and end-products (DMAPP and IPP) concentrations are, in general, larger than those of intermediate metabolites, which is another hallmark of a well-behaved biosynthetic pathway (Alves and Savageau, 2000).

To understand how perturbations in parameters may affect the ability of the system to maintain homeostasis, we calculated the logarithmic sensitivity of the steady state Jacobian eigenvalues to each parameter of the model (Supplementary Table S8). The model has over eighty parameters and eigenvalues have sensitivities that are above one (in absolute values) to thirty of them. The parameters to which more eigenvalues are sensitive concentrate in reactions r2 ($HMG\text{-}CoA_{cyt} \rightarrow MVA_{cyt}$), r3 ($MVA_{cyt} \rightarrow MVP_{cyt}$), r4 ($MVP_{cyt} \rightarrow MVPP_{cyt}$), and r6 ($IPP_{cyt} \rightarrow DMAPP_{cyt}$) of the MVA pathway and reactions r10 (Glyceraldehyde-3-P + Pyruvate → DXP) and r18 ($IPP_{pl} \rightarrow DMAPP_{pl}$) of the MEP pathways. This suggests that modifying isoprenoid biosynthesis could destabilize the physiological steady states of the plant.

TABLE 1 Concentrations of the basal model.

Metabolites	[mM]
$HMGCoA_{cyt}$	0.983
MVA_{cyt}	$3.5x10^{-5}$
MVP_{cyt}	$3.98x10^{-4}$
$MVPP_{cyt}$	$3.36x10^{-5}$
IPP_{cyt}	0.109
IPP_{pla}	0.0801
$DMAPP_{cyt}$	0.136
$DMAPP_{pla}$	0.124
DXP	0.0133
MEP	$1.15x10^{-3}$
CDPME	$1.11x10^{-4}$
CDPMEP	0.0920
MECPP	0.657
HMBPP	$3.52x10^{-4}$

3.3 Homeostatic concentrations are robust to enzyme mutations in the WT line

Sensitivity analysis identifies the parameters to which the various variables of the model are most sensitive, as described in

TABLE 2 Eigenvalues for the Steady State.

	Real	Im
Eigenvalue1	-928.693	0
Eigenvalue2	-891.953	0
Eigenvalue3	-872.862	0
Eigenvalue4	-272.249	0
Eigenvalue5	-83.721	0
Eigenvalue6	-78.084	0
Eigenvalue7	-16.974	0
Eigenvalue8	-11.845	0
Eigenvalue9	-6.668	0
Eigenvalue10	-2.011	0
Eigenvalue11	-0.913	0
Eigenvalue12	-0.545	0
Eigenvalue13	-0.110	0
Eigenvalue14	-0.031	0

(Sorribas et al., 2007; Alves et al., 2008). A high sensitivity of a variable to a parameter indicates that small changes in that parameter might lead to big changes in the variable.

Plausible models of biological systems have low sensitivities to most parameters (Savageau, 1976; Kitano, 2007). The logarithmic sensitivity analysis of the dependent concentrations of the WT model with respect to each parameter of the model we performed shows that our model fits this quality criterion. Only 51 out of 728 sensitivities are larger than one (Supplementary Table S9). DMAPP and IPP are the metabolites with the highest sensitivities. High sensitivities are well known to identify the parts of a system that need to be modeled in more detail when additional information becomes available (Savageau, 1976; Kitano, 2007). This is consistent with the fact that we modeled IPP and DMAPP usage only through simple sink reactions, without considering any metabolic and regulatory details.

3.4 Existence of homeostatic behavior in the mutant lines requires posttranscriptional regulation of protein activity

We implement the models for the biosynthetic pathways in each mutant line using the same procedure as that for the WT line (sections 2.3 to 2.6). These systems do not reach homeostasis, having unstable steady states with a few intermediate metabolites accumulating over time. This suggests that the models are not plausible representations of the biological situation (Savageau, 1971; Savageau, 1975; Savageau, 1976; Voit, 2013).

Biological systems can stabilize steady states and reach homeostasis by adjusting the activity of enzymes in a pathway, for example through post-transcriptional regulation of protein levels and activity. We investigated whether emulating this type of adjustment would stabilize the steady states in the models.

First, we identified the metabolites that accumulated in each line, which were DXP, CDP-MEP, MEcPP, or combinations thereof. Reactions r11, r14, and r15 of Table S2 either produce or consume these metabolites. Using a minimal intervention policy, we scanned the values for the Vmax parameters of reactions Vmax9, Vmax12 and Vmax13 in order to identify the minimum change in those parameters that would stabilize the steady state of each mutant line.

To stabilize homeostasis in the models we scanned one-dimensional, two-dimensional and three-dimensional parameter spaces and found the values of Vmax that stabilized the steady state in each model. The sets of Vmax that made the model for each mutant line stable were stored in a candidate sets list and we chose the final parameter set as the one with minimum normalized Euclidian distance to the original set of parameter values for that line. We provide the list of stabilized parameter values for each line in Supplementary Data S1 – Model_stabilization. Supplementary Figure S1 shows that Type II and Type III mutants require larger changes in parameter values than Type I mutants.

3.5 Stabilized homeostatic concentrations are robust to enzyme mutations in the mutant lines

We performed a sensitivity analysis of the stable homeostatic concentrations with respect to each enzyme parameter, in each mutant line (Figure 2A). We find that those concentrations are very robust, with 3% of all individual sensitivities being higher than 1 in absolute value in Type I mutants. This number goes down to 1.8% in Type II mutants and 1.4% in Type III mutants. The total number of individual sensitivities calculated for each line is higher than 900.

Globally, we found that, for each parameter p_j, the aggregated sensitivity S'_j of all metabolites to that parameter decreases in mutant lines with respect to the WT. The global sensitivities to each parameter decrease in the following order: WT>Type I lines> Type II lines> Type III lines. We also found that the aggregated sensitivity of each metabolite M_i to all parameters also decreases in the same order (Figure 2B). Thus, our results suggest that post-transcriptional regulation of a small number of enzymes is sufficient to maintain homeostasis of IPP/DMAPP biosynthesis in each mutant line.

3.6 Investigating average behavior for each mutant type

While having a line-specific, data-driven, model is a more accurate way of describing and predicting the behavior of each mutant line, these are less than helpful in predicting how a new mutant line of any of the three types will behave dynamically.

To create general, type-specific models that are more useful for predicting the dynamic behavior and characteristics of generic new mutant lines, we created a median experimental line for each mutant type. To do so we use the median gene expression activities for each gene in all lines of a given mutant type. Then, we follow the procedure described in methods to generate three new models, one per mutant type. Their steady state concentrations, stability, and sensitivity analysis in Supplementary Table S10; Figure 2; Supplementary Data S1. Figure 2 also shows that these lines have sensitivity profiles that are similar to those of the individual mutant lines of the same type. Moreover, the models for the median lines of each mutant type have homeostatic behavior that is robust to mutations in enzyme parameters, which is a hallmark of a plausible model.

3.7 Variations in whole plant hormone levels partially explain variations in the biosynthesis of IPP/DMAPP in seeds

We also wanted to investigate whether early plant hormone levels in the plant might be a proxy for subsequent changes in the biosynthesis of IPP/DMAPP in seeds. To do so, we calculated how changing hormone levels could explain changes in gene expression and metabolite levels, as described in the methods section 2.11 and

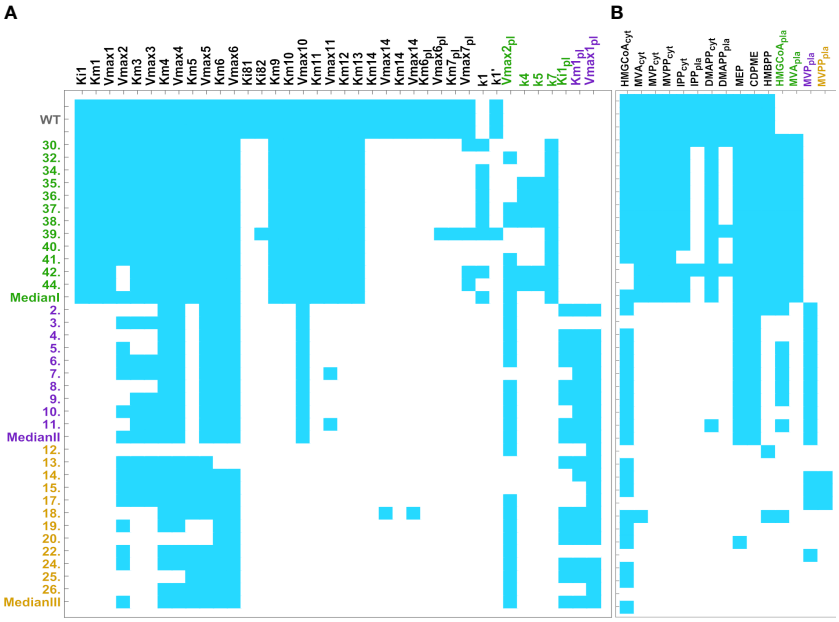


FIGURE 2
Heatmap of pooled sensitivities by parameter (left, threshold = 0.059) and by metabolite (right, threshold = 0.025), compared across mutant lines. Endogenous pathways are equal in all mutant types and have the same reactions. The models for the exogenous MVA plastid pathway are mutant type specific. Type III mutants include the reactions catalyzed by the five exogenous genes, HMGS, MVK, HMGR, PMK, and MVD. Type II mutants include the reactions catalyzed by HMGS, HMGR, and MVK. Type I mutants only include the reaction catalyzed by exogenous HMGR. As such, type III models have more kinetic parameters than type II. Type I has the lowest number of parameters. **(A)** Aggregate sensitivity of all metabolites to each parameter. **(B)** Aggregate sensitivity of each metabolite to all parameters.

in Supplementary Appendix S1. We are not assuming that the endogenous MEP, MVA, and exogenous MVA Pathways have an influence in hormone production, only measuring if there is a phenomenological correlation between variations in hormone levels and metabolites or levels of gene expression.

iP is a proxy for changes in enzyme activity in all three mutant types. For all mutant types, iP correlates to changes in the activity of the early MEP pathway steps. In addition, for type I mutants, iP also correlates to changes in the activity of MVA pathway early steps. ABA is also a proxy influence in the early steps of the MEP pathway for all mutant types. Other hormones have a mutant specific effect. Table S7 summarizes the results for all hormones and presents the hormone dependency equations for each mutant type.

To validate the resulting multilevel models, we investigated if they could reproduce the correlation between experimental hormone levels and model metabolites in the following way.

For each mutant type and metabolite whose concentration is significantly affected by a given hormone, we took the median model described in Section 3.6 and calculated the concentration of the various metabolites as a function of hormone levels. Then, we calculated the correlation in the simulation plot and compared that correlation to the one computed when we plot concentration of the same metabolite *vs* experimental hormone levels. We summarize the results of this analysis in Table 3. We find that the models maintain 48 out of 53 expected correlations between hormone levels and metabolite concentrations. This is consistent with the extended models being plausible multilevel descriptions of IPP/DMAPP biosynthesis in the three mutant types.

3.8 IPP/DMAPP production increases in mutants with a complete exogenous MVA pathway in the plastid

We estimated how the production of IPP/DMAPP changes across mutant lines and types, with the help of the models. Figure 3 summarizes the results. Overall, our models suggest that median metabolic flux going into the plastid's IPP/DMAPP producing pathways increases in the following way: Type III>Type II>Type I. We dissected the production rates of IPP by HDR, IDI and MVD (both endogenous and ectopic, where applicable), and the exchange from plastid to cytosol. We also find that the mean production flux for IPP increases from Type I to Type II to Type III.

3.9 Plant phenotype can be correlated to hormone and gene expression levels

We used linear regression to investigate if hormone, gene expression, and metabolite levels can predict macroscopic plant phenotypes, such as plant height, number of leaves, leaf width, leaf length and chlorophyll levels, as described in methods. Table 4 shows that the subset of phenotype parameters that can be predicted from molecular data is different for different mutant types and provides the best fit models for each phenotype parameter and mutant type. Leaf length and chlorophyll content can be predicted for types I, II, and III. Height can be predicted for types I and III. Leaf width can be predicted for types II and III. The

TABLE 3 Qualitative assessment of expected correlations between metabolites and hormones.

	Number of correlations assessed	Number of correlations matching expected effect	Ambiguous effect or none observed
Type I	7	4	3
Type II	19	19	0
Type III	27	24	2

number of leaves can be predicted for mutant types I and II. Variations in the levels of the metabolites, genes and hormones shown in Table 4 are the best predictors for the variations measured in the macroscopic phenotype. The metabolic levels of DXP, MEcPP, HMG-CoA are useful in predicting phenotypical characters of leaves in all mutant types. DMAPP levels are useful in predicting chlorophyll content in type III mutants. Table 4 also includes the adjusted R^2 for each model, which is a measure of the percentage of variation in the phenotype that can be explained by the model. Models explain between 40% and 77% of the phenotypical variability, depending on the specific phenotype being measured and the mutant type.

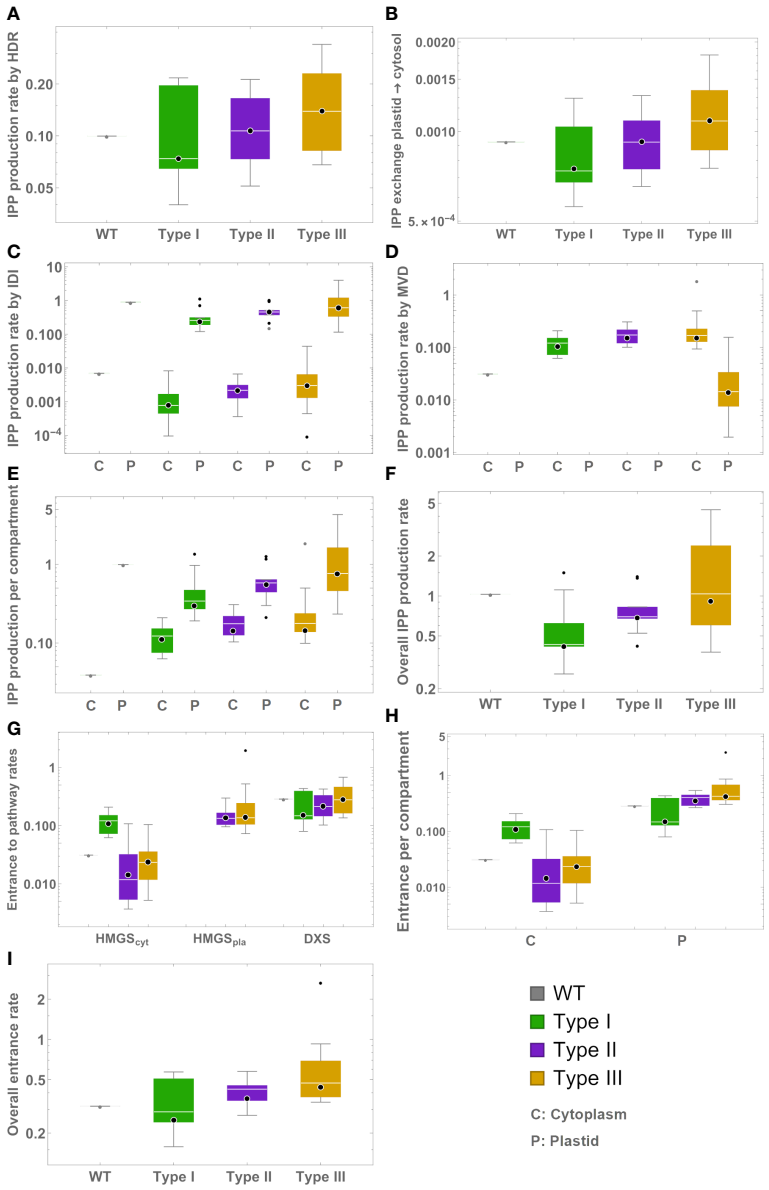


FIGURE 3
Box plots representing model predictions for exchange and production rates of IPP from different sources. Green – type I mutants. Mauve – type II mutants. Gold – type III mutants. C – Production in cytosol. P – Production in plastid. **(A)** Production by HDR. **(B)** Export to the cytosol. **(C)** Production by IDI **(D)** Production by MVD. **(E)** Total production in each compartment. **(F)** Total production. **(G)** Entry flux to the MVA, MEP, and exogenous MVA pathways. **(H)** Entry flux into IPP/DMAPP production per compartment. **(I)** Overall flux directed for IPP/DMAPP production.

TABLE 4 Multivariate linear models for the phenotype.

		Adjusted R^2
Type I	***Height*** = 56.028 – 22.688 *MDS* + 1.590 *DXP* ***Leaves*** = *Round* (4.738 + 0.318 *DXP* – 0.245 *MEcPP*) ***Leaf Length*** = 31.045 – 3.135 *HMGS* + 1549.320 *HMBPP* ***Chlorophyl*** = 32.218 + 0.154 ***IAA*** – 1.571 *HMGS*	**0.53** **0.72** **0.69** **0.74**
Type II	***Leaves*** = *Round* (3.325 – 0.00312 ***ACC*** + 0.0557 $HMGCoA_{cyt}$) ***Leaf Length*** = 41.621 + 6.048 *HMGS* – 116.641 *MEP* ***Leaf Width*** = 0.596 + 0.124 *HMGS* + 0.0464 *MVD1* ***Chlorophyl*** = 39.29 + 1.620 ***iP*** – 1.446 *HDR*	**0.70** **0.77** **0.61** **0.75**
Type III	***Height*** = 76.626 + 25.445 ***GA***4 + 0.432 *HMGS* ***Leaf Length*** = 50.826 + 0.197 *WR1* + 9071.25 $MVPP_{cyt}$ ***Leaf Width*** = 0.881 + 0.005 *HMGS* + 0.007 $HMGCoA_{cyt}$ ***Chlorophyl*** = 33.218 + 10.968 *HDR* + 75003 *DMAPP*	**0.60** **0.43** **0.40** **0.46**

All units in (cm), except number of Leaves. Blue variables indicate gene expression levels. Bold variables indicate hormone levels. All other variables represent metabolite levels.

3.10 Leveraging the models for phenotype prediction

The results from the previous section allow us to create a multilevel model, connecting metabolite concentrations, and hormone and gene expression levels to the macroscopic plant phenotypes. We summarize the multilevel model building process in Figure 4. In Figure 5 we use the models from Table 4 and the experimental data for gene expression and hormone levels to calculate what is the expected value for the predicted phenotype, according to the relevant model. Then, we include the experimental determination for the same phenotype. We show that each type-specific model can semi-quantitatively predict macroscopic plant phenotype.

3.11 Models are mutant-type specific

We further investigated if the type-specific models were accurate in predicting the phenotype of the other mutant types. We used the median model for each mutant type, feeding it with the experimental determinations for all the lines and measured how accurate the phenotype predictions were for each type. What we observed is that the type-specific models do not properly predict other mutant types. For example, Figure 6 shows that our type I model can only accurately predict chlorophyll levels for type I lines, failing to do so for mutant types II and III. Supplementary Table S11 shows that phenotype predictions are only accurate when made with the model for the correct mutant type. This suggests that prediction for new mutant types would require developing a data driven model for that mutant type.

4 Discussion

IPP and DMAPP are the precursor monomers to terpenoids, a family of molecules that contains many chemicals with importance in biology, pharmacy, biotechnology, biomedicine and cosmetics, such as squalene. Plants produce those monomers using two biosynthetic pathways: the MVA pathway in the cytosol, and the MEP pathway in the plastid. As such, they are in principle a good substrate for synthetic biology of valuable terpenoid biosynthesis. Plants engineered with exogenous MVA pathway genes are a

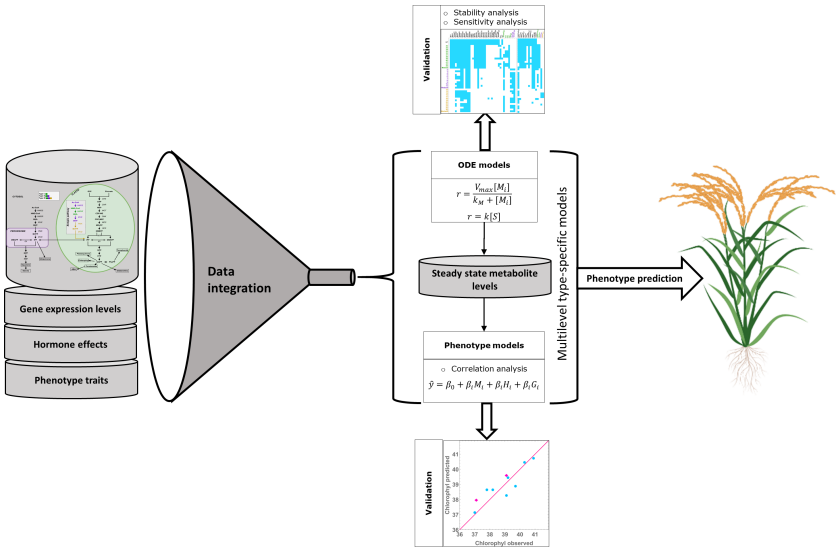


FIGURE 4
– Multilevel modeling process. We integrate pathway data with gene expression and hormone levels to create line-specific models for isoprenoid biosynthesis. We use the models to calculate metabolic steady state levels, which are then used as input variables, together with hormone levels, to model plant phenotype traits in a type specific manner. We validated the multilevel models and then used them to predict the phenotype of additional rice mutants of each type.

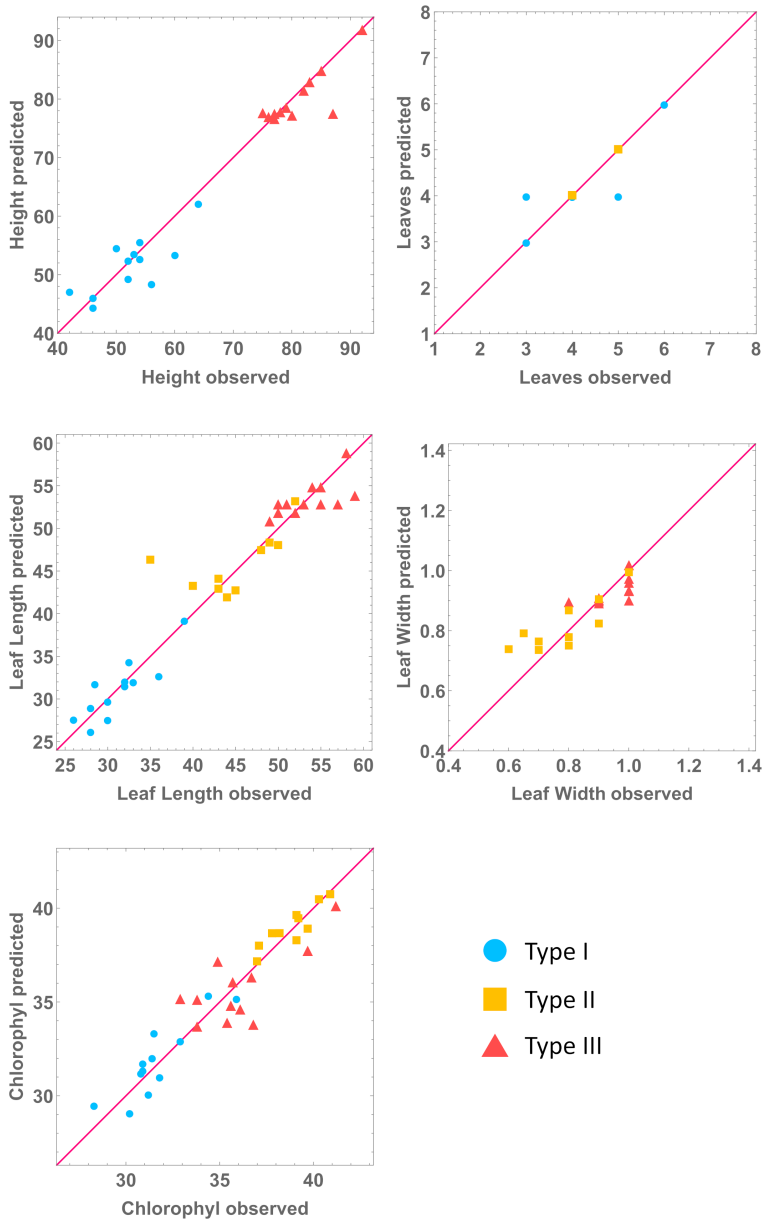


FIGURE 5
Predicted against observed values of phenotype variables. All units in cm, except chlorophyl (mg/g) and number of leaves. We predict phenotype for each mutant line by combining the molecular line-specific models with the phenotype, type-specific, models. Note: each type is being predicted by its own model.

promising platform for downstream terpenoid production (Andersen et al., 2021; Pérez et al., 2022). For example, when exogenous HMGRS, is expressed in the cytoplasm of tobacco leaves, these leaves appear to produce more cytoplasmic IPP/DMAPP. By also expressing exogenous crtE, crtB, and crtI, the plant uses the excess IPP/DMAPP to become biofortified in carotenoid pigments, such as lycopene (Andersen et al., 2021). Changing the flux going through the MEP plastid pathway has stronger, pleiotropic effects (Pérez et al., 2022). This is likely due to the developmental plant hormones produced from plastid IPP/DMAPP. As such, increasing the production of IPP/DMAPP in the chloroplast while containing the deleterious effects this might have in plant development is more effectively done through expression of an orthogonal MVA pathway in the plastid (Pérez et al., 2022). Still, plants use their IPP and DMAPP to synthesize all their cognate isoprenoids, including developmental hormones and protective molecules. Because of this, engineering plants to divert material from the pathways towards biotechnological purposes has significant pleiotropic effects that are often deleterious (Ye et al., 2000; Ro et al., 2006; Naqvi et al., 2009; Ajikumar et al., 2010; Zorrilla-López et al., 2013; Fuentes et al., 2016; Jiang et al., 2017; Kang et al., 2017; Georgiev et al., 2018; Kotopka et al., 2018; Cravens

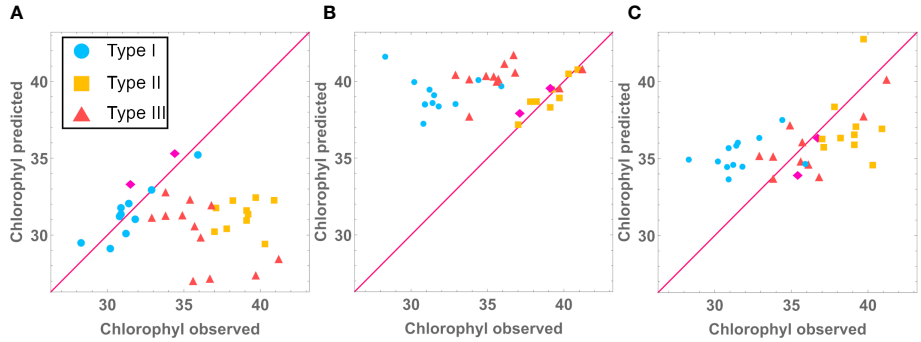


FIGURE 6
Using Type I **(A)**, Type II **(B)** and Type III **(C)** Chlorophyl models to predict Type I, Type II and Type III chlorophyl levels (mg/g).

et al., 2019; Luo et al., 2019; Maeda, 2019; Gülck et al., 2020; Srinivasan and Smolke, 2020; Zhu et al., 2021; Liew et al., 2022; Pérez et al., 2022; Yang et al., 2022; Zhang et al., 2022; Grzech et al., 2023). Consequently, if plants are to be used as a platform for terpenoid biosynthesis, one must engineer IPP/DMAPP biosynthesis in such a way that they are able to properly develop, while producing an excess of monomers that can be used for downstream high value terpenoids production.

This work aims to contribute towards that goal. We generated over thirty independent mutant rice lines that, in addition to the native MVA and MEP pathways, had three alternative versions of an exogenous MVA pathway located to the chloroplast. The method used to create the lines results in a non-targeted integration of the ectopic genes in the genome of the endosperms. This created the potential for different dynamic behavior of pathway metabolites among lines within the same mutant type. We then measured the expression of the genes in the pathways, the hormone levels, and the macroscopic phenotype of the WT and mutant lines. We combined all this data into multiscale, line-specific, mathematical models of the plants that connected all the variables and measurements. We use these models to understand how the alternative versions of the pathways contribute to change the flux going through the IPP/DMAPP metabolic pools.

4.1 Modelling limitations

Several modeling efforts focused on analyzing the biosynthetic and signaling dynamics of complex terpenoids in plants (Latowski et al., 2000; Bruggeman et al., 2001; Rios-Estepa et al., 2008; Rios-Estepa et al., 2010; Band et al., 2012; Beguerisse-Diaz et al., 2012; Pokhilko et al., 2013; Pokhilko et al., 2015; Allen and Ptashnyk, 2017; Nazareno and Hernandez, 2017; Dalwadi et al., 2018; Rizza et al., 2021). For example, Band et al. and Rizza et al. use compartmental modeling to study the biosynthesis and diffusion of gibberellins in root tips (Band et al., 2012; Rizza et al., 2021). Terpenoid signaling is often also a target for modeling. For example, Allen & Ptashnyk use models to study signaling interactions between brassinosteroid and gibberellin signaling pathways (Allen and Ptashnyk, 2017), and Nazareno & Hernandez do the same to study signaling interactions between of abscisic acid, ethylene and methyl jasmonate on stomatal closure (Nazareno and Hernandez, 2017). Modeling studies of the biosynthesis and regulation of terpenoid precursors in plants are less common. In fact, we are only aware of two such modeling effort for the MEP pathway in plants (Pokhilko et al., 2015) (Neiburga et al., 2023) and another in *Plasmodium falciparum* (Singh and Ghosh, 2013). The plant MEP model was used to study how circadian rhythms regulate the dynamics of the pathway in plants, while the *P. falciparum* model was used to investigate the regulation of the pathway and to predict the effects of genetic manipulations on the production of isoprenoids with the addition of *in silico* inhibitors. Regarding the MVA pathway, we found no model in plants. Still, this pathway was modeled in bacteria (Weaver et al., 2015, Dalwadi et al., 2018). These MVA models study the dynamics of the pathway in the context of introducing it in bacteria using synthetic biology. Finally, we know of only one other example where both pathways were modeled together in yeast, using Petri net-based dynamic modeling (Baadhe et al., 2012). Taking all this into account, creating models that can be used to study the dynamics and interactions of the MVA and MEP pathways in plants is an important goal, towards which this paper contributes.

Trusting the models normally requires that they are validated by comparison with experimental data that was not used to build them. This was one of the technical limitations in building line-specific models, as we needed most of the available data to estimate the parameter values for each line. As such, the measurements can be used for either model building or model validation, but not both simultaneously. To sidestep this problem, we used three approaches.

First, we used sensitivity and stability analysis as general model quality assessment tools (Savageau, 1975; Kitano, 2007; Voit, 2013) to both evaluate the quality of the line-specific models and identify the parameters that could be used to improve that quality. The WT model is of high quality, being robust to changes in parameters (>92% of parameter sensitivities lower than 0.5) and producing a stable steady state, with metabolite concentrations that are well within the accepted biological ranges (Albe et al., 1990). In contrast, the steady states for the original line specific mutant lines are unstable, leading to unbound accumulation of MEP pathway

intermediates. This indicated that the models needed improvement. We hypothesized that the simplest reason for model instability could be consequence of a nonlinear relationship between changes in the expression of pathway genes and changes in enzyme activity. To test if the hypothesis is consistent with our data, we scanned the parameters that represent the enzyme activities that directly produce or use the accumulating metabolites in the model: Vmax9, Vmax12 and Vmax 13. By executing this procedure, we identified the sets of minimal changes to the values for these parameters that generated stable and robust steady states, with metabolic concentrations within known physiological ranges. We remark that other, more complex explanations might also be consistent with the experimental data. Still, Occam's razor argues that the simplest explanation is the most likely one, in the absence of additional data (Borgqvist and Palmer, 2022; Piasini et al., 2023).

Second, we created "median" models for each of the three types of pathways. To do so, we pooled together all mutant lines of a given type and calculated the median of the pool for each variable. Then, we used that median to create the median model for each mutant type, in the same way we create line specific models. These median models also allowed us to estimate the metabolites for each individual line. Comparing the results with the line specific models shows that these median models have a similar behavior to the models of the individual lines of the same mutant type (Figure 2). In addition, when we use the median models to predict the phenotype of the individual lines, the predictions have an error that is similar to that of the individual line models. This suggests that we can use the median model of a given mutant type to study newly created lines for that mutant type.

Third, we used the multilevel models to predict plant phenotype and compare the results with the experimentally determined phenotype, achieving a prediction accuracy of up to 80% (Supplementary Table S11). We summarize the process in Figure 4. We make different assumptions about the relationship between variables in our modeling. We model the effect of changes in gene expression on the concentrations of pathway intermediates assuming a direct cause-effect relationship between changes in the expression of a gene and changes in the corresponding enzyme activity. We model the effect of enzyme activity on the flux going through the reaction catalyzed by that enzyme using traditional enzyme kinetics. In both cases this assumes a causal relationship between variables. In contrast, we assume that there might be a phenomenological relationship between changes in hormone levels and changes in gene expression and phenotypes and test for that relationship. When we find statistical evidence for that relationship, we include it in our models in different ways. The influence of hormone levels on gene expression is added to the metabolic pathway models through the use of approximation theory, in a way that is mathematically well justified (Salvador, 1996; Sorribas et al., 2007; Alves et al., 2008). The phenomenological influence of hormones, genes, and metabolites on phenotype was accounted for by using statistical linear models. As additional experimental studies become available, the phenomenological parts of the models will need to be adjusted in order to account for the knowledge generated by those experiments. Overall, the quality assessment steps we performed suggest that our models can be used as reasonable semi-quantitative prediction tools to help in better understanding the biology of isoprenoid biosynthesis modification in rice.

4.2 Biology of IPP/DMAPP production: from molecular determinants to plant phenotype

Posttranscriptional regulation is important for the proper functioning of the MEP and MVA pathways in plants (Laule et al., 2003; Guevara-García et al., 2005; Sauret-Güeto et al., 2006; Flores-Pérez et al., 2008; Xie et al., 2008; Cordoba et al., 2009; Han et al., 2013). Our modeling and analysis indicate that IPP/DMAPP production is robust to fluctuations in enzyme activity in WT rice (Table 2; Supplementary Tables S8, S9). Further, it suggests that posttranscriptional modulation of enzyme activity is important in stabilizing IPP/DMAPP production in mutant lines (Figure 2; Supplementary Table S2). In all cases, stabilizing the steady state of a mutant line requires that the activity of a protein is upregulated with respect to the changes in gene expression for that protein (Supplementary Data S1). This is fully consistent with the observation that, when compared to the WT *Arabidopsis* plants, changes in the activities of DXS and DXR proteins are bigger than the changes in expression of the corresponding genes (compare panels A and C of Figure 3 in Flores-Pérez et al. (2008)). Interestingly, DXR is one of the proteins flagged in our models as a potential stabilizing influence for the steady state of the rice mutants (Supplementary Data S1).

Our results also suggest that variations in plant hormone levels can predict, to some extent, both plant phenotype (Table S11) and the biosynthetic fluxes of IPP/DMAPP in the seeds (Table S7). Further, variations in plant gene expression levels in combination with variations in plant hormones can improve phenotype predictions (Tables 4, S11). However, the more complex the genetic manipulation was, the less accurate the phenotype predictions become. While the median adjusted R^2 for the predictions is approximately 60% in mutant types I and I, this number goes down to 35% in type III mutants (Table S11).

Pérez et al. (2022) also reported that plant development is more similar to that of the WT in mutant type III, followed by mutant type II, and finally I. Our modeling suggests an explanation for this observation. The analysis predicts that the global production of IPP/DMAPP is on average higher in the mutant types that have a more complete version of the exogenous MVA pathway in the plastid (Figure 3). The average amount of flux going through the endogenous MEP and MVA pathways in Type III mutants is the most similar to that of the WT, followed by the flux going through the endogenous pathways in Type II mutants. The least similar flux to WT is that of Type I mutants (Figure 3). Figure 3 also shows that the total amount of flux going through the IPP/DMAPP pools in mutant types I and II is similar to that of the WT. The flux going through IPP/DMAPP is larger in Type III mutants than in the WT rice. Taken together, these observations suggest that plants can distinguish between the IPP/DMAPP produced by each of the pathways. Too little flux going through the endogenous MEP

pathway compromises the production of developmental hormones leading to plants with developmental defects. In conclusion, we believe that an iterative modeling-experimental process as the one presented here would be an effective way to identify which parts of each pathway are more sensitive to further manipulation, and which parts are more likely to be good targets for modification in order to increase the flux without disrupting the development of the plant.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Author contributions

OB and RA designed the modelling and performed the in silico analysis. LP, PC and TC designed the plasmids, transformed and recovered the plants, and measured phenotypes. LF and PF designed and performed the metabolomics experiments. AA and CA designed and performed the hormone measurements. OB, RA, AL, EV, AM-S and AS analyzed results. OB and RA wrote the paper. All authors contributed to the article and approved the submitted version.

Funding

PROSTRIG, an ERANET project from FACEJPI (PCI2019-103382, MICIUN), partially funded this project. AL received funding from the European Union's H2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No. 801586. OB received a Ph. D. fellowship from AGAUR (2022FI_B 00395). LP received a Ph. D. fellowship from MINECO. This work was also partially supported by The Spanish Ministry of Economy and Competitiveness (MINECO, project BIO2014-54426-P) and the European Union Framework Program DISCO (from DISCOvery to products: a next-generation pipeline for the sustainable generation of high-value plant products, project 613513) also partially funded this work.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1133299/full#supplementary-material

References

Ajikumar, P. K., Xiao, W. H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., et al. (2010). Isoprenoid pathway optimization for taxol precursor overproduction in escherichia coli. *Science* 330, 70–74. doi: 10.1126/SCIENCE.1191652

Albe, K. R., Butler, M. H., and Wright, B. E. (1990). Cellular concentrations of enzymes and their substrates. *J. Theor. Biol.* 143, 163–195. doi: 10.1016/S0022-5193(05)80266-8

Allen, H. R., and Ptashnyk, M. (2017). Mathematical modelling and analysis of the brassinosteroid and gibberellin signalling pathways and their interactions. *J. Theor. Biol.* 432, 109–131. doi: 10.1016/J.JTBI.2017.08.013

Alves, R., and Savageau, M. A. (2000). Comparing systemic properties of ensembles of biological networks by graphical and statistical methods. *Bioinformatics* 16, 527–533. doi: 10.1093/BIOINFORMATICS/16.6.527

Alves, R., Vilaprinyo, E., Hernádez-Bermejo, B., and Sorribas, A. (2008). Mathematical formalisms based on approximated kinetic representations for modeling genetic and metabolic pathways. *Biotechnol. Genet. Eng. Rev.* 25, 1–40. doi: 10.5661/bger-25-1

Andersen, T. B., Llorente, B., Morelli, L., Torres-Montilla, S., Bordanaba-Florit, G., Espinosa, F. A., et al. (2021). An engineered extraplastidial pathway for carotenoid biofortification of leaves. *Plant Biotechnol. J.* 19, 1008–1021. doi: 10.1111/PBI.13526

Baadhe, R. R., Mekala, N. K., Palagiri, S. R., and Parcha, S. R. (2012). Development of petri net-based dynamic model for improved production of farnesyl pyrophosphate by integrating mevalonate and methylerythritol phosphate pathways in yeast. *Appl. Biochem. Biotechnol.* 167, 1172–1182. doi: 10.1007/S12010-012-9583-1/FIGURES/9

Band, L. R., Úbeda-Tomás, S., Dyson, R. J., Middleton, A. M., Hodgman, T. C., Owen, M. R., et al. (2012). Growth-induced hormone dilution can explain the dynamics of plant root cell elongation. *Proc. Natl. Acad. Sci. USA* 109, 7577–7582. doi: 10.1073/PNAS.1113632109/SUPPL_FILE/SAPP.PDF

Beguerisse-Diaz, M., Hernández-Gómez, M. C., Lizzul, A. M., Barahona, M., and Desikan, R. (2012). Compound stress response in stomatal closure: a mathematical model of ABA and ethylene interaction in guard cells. *BMC Syst. Biol.* 6, 1–15. doi: 10.1186/1752-0509-6-146/FIGURES/7

Bick, J. A., and Lange, B. M. (2003). Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane. *Arch. Biochem. Biophys.* 415, 146–154. doi: 10.1016/S0003-9861(03)00233-9

Borgqvist, J. G., and Palmer, S. (2022). Occam's razor gets a new edge: the use of symmetries in model selection. *J. R. Soc Interface* 19, 20220324. doi: 10.1098/RSIF.2022.0324

Bruggeman, F. J., Libbenga, K. R., and Van Duijn, B. (2001). The diffusive transport of gibberellins and abscisic acid through the aleurone layer of germinating barley grain: a mathematical model. *Planta* 214, 89–96. doi: 10.1007/S004250100588/METRICS

Chen, J., Tan, J., Duan, X., Wang, Y., Wen, J., Li, W., et al. (2023). Plastidial engineering with coupled farnesyl diphosphate pool reconstitution and enhancement for sesquiterpene biosynthesis in tomato fruit. *Metab. Eng.* 77, 41–52. doi: 10.1016/J.YMBEN.2023.03.002

Comas, J., Benfeitas, R., Vilaprinyo, E., Sorribas, A., Solsona, F., Farré, G., et al. (2016). Identification of line-specific strategies for improving carotenoid production in synthetic maize through data-driven mathematical modeling. *Plant J.* 87, 455–471. doi: 10.1111/TPJ.13210

Cordoba, E., Salmi, M., and León, P. (2009). Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants. *J. Exp. Bot.* 60, 2933–2943. doi: 10.1093/JXB/ERP190

Cravens, A., Payne, J., and Smolke, C. D. (2019). Synthetic biology strategies for microbial biosynthesis of plant natural products. *Nat. Commun.* 10, 1–12. doi: 10.1038/s41467-019-09848-w

Dalwadi, M. P., Garavaglia, M., Webb, J. P., King, J. R., and Minton, N. P. (2018). Applying asymptotic methods to synthetic biology: modelling the reaction kinetics of the mevalonate pathway. *J. Theor. Biol.* 439, 39–49. doi: 10.1016/J.JTBI.2017.11.022

Dudareva, N., Andersson, S., Orlova, I., Gatto, N., Reichelt, M., Rhodes, D., et al. (2005). The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers. *Proc. Natl. Acad. Sci. USA* 102, 933–938. doi: 10.1073/PNAS.0407360102/ASSET/F5C94EBD-80FE-4A87-A155-E7CC4E980CE3/ASSETS/GRAPHIC/ZPQ0030569670005.JPEG

Efroymson, M. A. (1960). "Multiple regression analysis," in *Mathematical methods for digital computers*. Eds. A. Ralston and H. S. Wilf (New York: John Wiley).

Eisenreich, W., Rohdich, F., and Bacher, A. (2001). Deoxyxylulose phosphate pathway to terpenoids. *Trends Plant Sci.* 6, 78–84. doi: 10.1016/S1360-1385(00)01812-4

Flores-Pérez, Ú., Sauret-Güeto, S., Gas, E., Jarvis, P., and Rodríguez-Concepcióna, M. (2008). A mutant impaired in the production of plastome-encoded proteins uncovers a mechanism for the homeostasis of isoprenoid biosynthetic enzymes in arabidopsis plastids. *Plant Cell* 20, 1303–1315. doi: 10.1105/TPC.108.058768

Fuentes, P., Zhou, F., Erban, A., Karcher, D., Kopka, J., and Bock, R. (2016). A new synthetic biology approach allows transfer of an entire metabolic pathway from a medicinal plant to a biomass crop. *Elife* 5, e13664. doi: 10.7554/ELIFE.13664

Georgiev, V., Slavov, A., Vasileva, I., and Pavlov, A. (2018). Plant cell culture as emerging technology for production of active cosmetic ingredients. *Eng. Life Sci.* 18, 779–798. doi: 10.1002/ELSC.201800066

Grimbs, S., Selbig, J., Bulik, S., Holzhütter, H. G., and Steuer, R. (2007). The stability and robustness of metabolic states: identifying stabilizing sites in metabolic networks. *Mol. Syst. Biol.* 3, 146. doi: 10.1038/MSB4100186

Grzech, D., Hong, B., Caputi, L., Sonawane, P. D., and O'Connor, S. E. (2023). Engineering the biosynthesis of late-stage vinblastine precursors precondylocarpine acetate, catharanthine, tabersonine in nicotiana benthamiana. *ACS Synth. Biol.* 12, 27–34. doi: 10.1021/ACSSYNBIO.2C00434/SUPPL_FILE/SB2C00434_SI_001.PDF

Guevara-García, A., San Román, C., Arroyo, A., Cortés, M. E., de la Gutiérrez-Nava, M. L., and León, P. (2005). Characterization of the arabidopsis clb6 mutant illustrates the importance of posttranscriptional regulation of the methyl-d-Erythritol 4-phosphate pathway. *Plant Cell* 17, 628–643. doi: 10.1105/TPC.104.028860

Gülck, T., Booth, J. K., Khakimov, B., Crocoll, C., Motawia, M. S., Møller, B. L., et al. (2020). Synthetic biology of cannabinoids and cannabinoid glucosides in nicotiana benthamiana and saccharomyces cerevisiae. *J. Nat. Prod.* 83, 2877–2893. doi: 10.1021/ACS.JNATPROD.0C00241/ASSET/IMAGES/LARGE/NP0C00241_0009.JPEG

Han, M., Heppel, S. C., Su, T., Bogs, J., Zu, Y., An, Z., et al. (2013). Enzyme inhibitor studies reveal complex control of methyl-d-erythritol 4-phosphate (MEP) pathway enzyme expression in catharanthus roseus. *PloS One* 8 (5), e62467. doi: 10.1371/JOURNAL.PONE.0062467

Harborne, J. B., Tomas-Barberan, F. A.Phytochemical Society of Europe (1991) *Ecological chemistry and biochemistry of plant terpenoids* (Clarendon Press). Available at: http://agris.fao.org/agris-search/search.do?recordID=US201300687965 (Accessed June 4, 2019).

Hemmerlin, A. (2013). Post-translational events and modifications regulating plant enzymes involved in isoprenoid precursor biosynthesis. *Plant Sci.* 203–204, 41–54. doi: 10.1016/J.PLANTSCI.2012.12.008

Hemmerlin, A., Harwood, J. L., and Bach, T. J. (2012). A raison d'être for two distinct pathways in the early steps of plant isoprenoid biosynthesis? *Prog. Lipid Res.* 51, 95–148. doi: 10.1016/J.PLIPRES.2011.12.001

Hemmerlin, A., Hoeffler, J. F., Meyer, O., Tritsch, D., Kagan, I. A., Grosdemange-Billiard, C., et al. (2003). Cross-talk between the cytosolic mevalonate and the plastidial methylerythritol phosphate pathways in tobacco bright yellow-2 cells. *J. Biol. Chem.* 278, 26666–26676. doi: 10.1074/jbc.M302526200

Hemmerlin, A., Tritsch, D., Hartmann, M., Pacaud, K., Hoeffler, J. F., Van Dorsselaer, A., et al. (2006). A cytosolic arabidopsis d-xylulose kinase catalyzes the phosphorylation of 1-Deoxy-d-Xylulose into a precursor of the plastidial isoprenoid pathway. *Plant Physiol.* 142, 441–457. doi: 10.1104/PP.106.086652

Jiang, G. Z., Yao, M. D., Wang, Y., Zhou, L., Song, T. Q., Liu, H., et al. (2017). Manipulation of GES and ERG20 for geraniol overproduction in saccharomyces cerevisiae. *Metab. Eng.* 41, 57–66. doi: 10.1016/J.YMBEN.2017.03.005

Kang, A., Meadows, C. W., Canu, N., Keasling, J. D., and Lee, T. S. (2017). High-throughput enzyme screening platform for the IPP-bypass mevalonate pathway for isopentenol production. *Metab. Eng.* 41, 125–134. doi: 10.1016/J.YMBEN.2017.03.010

Kitano, H. (2007). Towards a theory of biological robustness. *Mol. Syst. Biol.* 3, 137. doi: 10.1038/MSB4100179

Kotopka, B. J., Li, Y., and Smolke, C. D. (2018). Synthetic biology strategies toward heterologous phytochemical production. *Nat. Prod. Rep.* 35, 902–920. doi: 10.1039/C8NP00028J

Kumar, S., Hahn, F. M., Baidoo, E., Kahlon, T. S., Wood, D. F., McMahan, C. M., et al. (2012). Remodeling the isoprenoid pathway in tobacco by expressing the cytoplasmic mevalonate pathway in chloroplasts. *Metab. Eng.* 14, 19–28. doi: 10.1016/J.YMBEN.2011.11.005

Lange, I., Poirier, B. C., Herron, B. K., and Lange, B. M. (2015). Comprehensive assessment of transcriptional regulation facilitates metabolic engineering of isoprenoid accumulation in arabidopsis. *Plant Physiol.* 169, 1595–1606. doi: 10.1104/PP.15.00573

Latowski, D., Burda, K., and Strzałka, K. (2000). A mathematical model describing kinetics of conversion of violaxanthin to zeaxanthin *via* intermediate antheraxanthin by the xanthophyll cycle enzyme violaxanthin de-epoxidase. *J. Theor. Biol.* 206, 507–514. doi: 10.1006/JTBI.2000.2141

Laule, O., Fürholz, A., Chang, H.-S., Zhu, T., Wang, X., Heifetz, P. B., et al. (2003). Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in arabidopsis thaliana. *Proc. Natl. Acad. Sci. USA* 100, 6866–6871. doi: 10.1073/pnas.1031755100

Liao, P., Hemmerlin, A., Bach, T. J., and Chye, M.-L. (2016). The potential of the mevalonate pathway for enhanced isoprenoid production. *Biotechnol. Adv.* 34, 697–713. doi: 10.1016/J.BIOTECHADV.2016.03.005

Liew, F. E., Nogle, R., Abdalla, T., Rasor, B. J., Canter, C., Jensen, R. O., et al. (2022). Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale. *Nat. Biotechnol.* 40 (3), 335–344. doi: 10.1038/s41587-021-01195-w

Luo, X., Reiter, M. A., d'Espaux, L., Wong, J., Denby, C. M., Lechner, A., et al. (2019). Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nat.* 567, 123–126. doi: 10.1038/s41586-019-0978-9

Maeda, H. A. (2019). Harnessing evolutionary diversification of primary metabolism for plant synthetic biology. *J. Biol. Chem.* 294, 16549–16566. doi: 10.1074/JBC.REV119.006132

Mcgarvey, D. J., and Croteau, R. (1995) *Terpenoid metabolism. American society of plant physiologists*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC160903/pdf/071015.pdf (Accessed June 5, 2019).

Naqvi, S., Zhu, C., Farre, G., Ramessar, K., Bassie, L., Breitenbach, J., et al. (2009). Transgenic multivitamin corn through biofortification of endosperm with three vitamins representing three distinct metabolic pathways. *Proc. Natl. Acad. Sci. USA* 106, 7762–7767. doi: 10.1073/PNAS.0901412106/SUPPL_FILE/0901412106SI.PDF

Navale, G. R., Dharne, M. S., and Shinde, S. S. (2021). Metabolic engineering and synthetic biology for isoprenoid production in escherichia coli and saccharomyces cerevisiae. *Appl. Microbiol. Biotechnol.* 105 (2), 457–475. doi: 10.1007/S00253-020-11040-W

Nazareno, A. L., and Hernandez, B. S. (2017). A mathematical model of the interaction of abscisic acid, ethylene and methyl jasmonate on stomatal closure in plants. *PloS One* 12, e0171065. doi: 10.1371/JOURNAL.PONE.0171065

Neiburga, K. D., Muiznieks, R., Zake, D. M., Pentjuss, A., Komasilovs, V., Rohwer, J., et al. (2023). Total optimization potential (TOP) approach based constrained design of isoprene and cis-abienol production in a. thaliana. *Biochem. Eng. J.* 190, 108723. doi: 10.1016/J.BEJ.2022.108723

Ni, T. C., and Savageau, M. A. (1996a). Application of biochemical systems theory to metabolism in human red blood cells. signal propagation and accuracy of representation. *J. Biol. Chem.* 271, 7927–7941. doi: 10.1074/JBC.271.14.7927

Ni, T. C., and Savageau, M. A. (1996b). Model assessment and refinement using strategies from biochemical systems theory: application to metabolism in human red blood cells. *J. Theor. Biol.* 179, 329–368. doi: 10.1006/JTBI.1996.0072

Page, J. E., Hause, G., Raschke, M., Gao, W., Schmidt, J., Zenk, M. H., et al. (2004). Functional analysis of the final steps of the 1-deoxy-d-xylulose 5-phosphate (DXP) pathway to isoprenoids in plants using virus-induced gene silencing. *Plant Physiol.* 134, 1401–1413. doi: 10.1104/PP.103.038133

Pérez, L., Alves, R., Perez-Fons, L., Albacete, A., Farré, G., Soto, E., et al. (2022). Multilevel interactions between native and ectopic isoprenoid pathways affect global metabolism in rice. *Transgenic Res.* 31, 249–268. doi: 10.1007/S11248-022-00299-6

Piasini, E., Liu, S., Chaudhari, P., Balasubramanian, V., and Gold, J. I. (2023). How occam's razor guides human decision-making. *bioRxiv*. doi: 10.1101/2023.01.10.523479

Pokhilko, A., Bou-Torrent, J., Pulido, P., Rodríguez-Concepción, M., and Ebenhöh, O. (2015). Mathematical modelling of the diurnal regulation of the MEP pathway in arabidopsis. *New Phytol.* 206, 1075–1085. doi: 10.1111/NPH.13258

Pokhilko, A., Mas, P., and Millar, A. J. (2013). Modelling the widespread effects of TOC1 signalling on the plant circadian clock and its outputs. *BMC Syst. Biol.* 7, 1–12. doi: 10.1186/1752-0509-7-23/FIGURES/6

Rios-Estepa, R., Lange, I., Lee, J. M., and Markus Lange, B. (2010). Mathematical modeling-guided evaluation of biochemical, developmental, environmental, and genotypic determinants of essential oil composition and yield in peppermint leaves. *Plant Physiol.* 152, 2105–2119. doi: 10.1104/PP.109.152256

Rios-Estepa, R., Turner, G. W., Lee, J. M., Croteau, R. B., and Lange, B. M. (2008). A systems biology approach identifies the biochemical mechanisms regulating monoterpenoid essential oil composition in peppermint. *Proc. Natl. Acad. Sci. USA* 105, 2818–2823. doi: 10.1073/PNAS.0712314105/SUPPL_FILE/12314APPENDIX.PDF

Rizza, A., Tang, B., Stanley, C. E., Grossmann, G., Owen, M. R., Band, L. R., et al. (2021). Differential biosynthesis and cellular permeability explain longitudinal

gibberellin gradients in growing roots. *Proc. Natl. Acad. Sci. USA* 118, e1921960118. doi: 10.1073/PNAS.1921960118/SUPPL_FILE/PNAS.1921960118.SM03.AVI

Ro, D. K., Paradise, E. M., Quellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440, 940–943. doi: 10.1038/NATURE04640

Salvador, A. (1996). Development of methodology and software for analysis of kinetic models of metabolic processes. application to the mitochondrial metabolism of lipid hydroperoxides. doi: 10.13140/RG.2.2.25894.40004

Sauret-Güeto, S., Botella-Pavía, P., Flores-Pérez, U., Martínez-García, J. F., San Román, C., León, P., et al. (2006). Plastid cues posttranscriptionally regulate the accumulation of key enzymes of the methylerythritol phosphate pathway in arabidopsis. *Plant Physiol.* 141, 75–84. doi: 10.1104/PP.106.079855

Savageau, M. A. (1971). Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature* 229, 542–544. doi: 10.1038/229542A0

Savageau, M. A. (1975). Significance of autogenously regulated and constitutive synthesis of regulatory proteins in repressible biosynthetic systems. *Nature* 258, 208–214. doi: 10.1038/258208A0

Savageau, M. A. (1976). *Biochemical systems analysis: a study of function and design in molecular biology* (United States: Reading, Mass, Addison-Wesley), Vol. 379.

Schaller, H., Grausem, B., Benveniste, P., Chye, M. L., Tan, C. T., Song, Y. H., et al. (1995). Expression of the hevea brasiliensis (H.B.K.) mull. arg. 3-Hydroxy-3-Methylglutaryl-Coenzyme a reductase 1 in tobacco results in sterol overproduction. *Plant Physiol.* 109, 761–770. doi: 10.1104/pp.109.3.761

Schmidt, M. D., Vallabhajosyula, R. R., Jenkins, J. W., Hood, J. E., Soni, A. S., Wikswo, J. P., et al. (2011). Automated refinement and inference of analytical models for metabolic networks. *Phys. Biol.* 8, 55011. doi: 10.1088/1478-3975/8/5/055011

Schneider, M. M., Hampp, R., and Ziegler, H. (1977). Envelope permeability to possible precursors of carotenoid biosynthesis during chloroplast-chromoplast transformation. *Plant Physiol.* 60, 518–520. doi: 10.1104/pp.60.4.518

Shih, P. M. (2018). Towards a sustainable bio-based economy: redirecting primary metabolism to new products with plant synthetic biology. *Plant Sci.* 273, 84–91. doi: 10.1016/J.PLANTSCI.2018.03.012

Singh, V. K., and Ghosh, I. (2013). Methylerythritol phosphate pathway to isoprenoids: kinetic modeling and in silico enzyme inhibitions in plasmodium falciparum. *FEBS Lett.* 587, 2806–2817. doi: 10.1016/J.FEBSLET.2013.06.024

Sorribas, A., Hernández-Bermejo, B., Vilaprinyo, E., and Alves, R. (2007). Cooperativity and saturation in biochemical networks: a saturable formalism using Taylor series approximations. *Biotechnol. Bioeng.* 97, 1259–1277. doi: 10.1002/bit.21316

Srinivasan, P., and Smolke, C. D. (2020). Biosynthesis of medicinal tropane alkaloids in yeast. *Nat.* 585 (7826), 614–619. doi: 10.1038/s41586-020-2650-9

Tetali, S. D. (2019). Terpenes and isoprenoids: a wealth of compounds for global use. *Planta* 249, 1–8. doi: 10.1007/S00425-018-3056-X/FIGURES/2

Voit, E. O. (1991). *Canonical nonlinear modeling: s-system approach to understanding complexity* (New york: Springer), Vol. 365.

Voit, E. O. (2013). Biochemical systems theory: a review. *ISRN Biomath.* 2013, 1–53. doi: 10.1155/2013/897658

Wang, R., Yin, Y., Li, J., Wang, H., Lv, W., Gao, Y., et al. (2022). Global stable-isotope tracing metabolomics reveals system-wide metabolic alternations in aging drosophila. *Nat. Commun. 2022* 13 (1), 1–14. doi: 10.1038/s41467-022-31268-6

Weaver, L. J., Sousa, M. M. L., Wang, G., Baidoo, E., Petzold, C. J., and Keasling, J. D. (2015). A kinetic-based approach to understanding heterologous mevalonate pathway function in e. coli. *Biotechnol. Bioeng.* 112, 111–119. doi: 10.1002/BIT.25323

Wright, L. P., Rohwer, J. M., Ghirardo, A., Hammerbacher, A., Ortiz-Alcaide, M., Raguschke, B., et al. (2014). Deoxyxylulose 5-phosphate synthase controls flux through the methylerythritol 4-phosphate pathway in arabidopsis. *Plant Physiol.* 165, 1488–1504. doi: 10.1104/PP.114.245191

Xie, Z., Kapteyn, J., and Gang, D. R. (2008). A systems biology investigation of the MEP/terpenoid and shikimate/phenylpropanoid pathways points to multiple levels of metabolic control in sweet basil glandular trichomes. *Plant J.* 54, 349–361. doi: 10.1111/J.1365-313X.2008.03429.X

Yang, Y., Chaffin, T. A., Ahkami, A. H., Blumwald, E., and Stewart, C. N. (2022). Plant synthetic biology innovations for biofuels and bioproducts. *Trends Biotechnol.* 40, 1454–1468. doi: 10.1016/J.TIBTECH.2022.09.007

Ye, X., Al-Babili, S., Klöti, A., Zhang, J., Lucca, P., Beyer, P., et al. (2000). Engineering the provitamin a (beta-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. *Science* 287, 303–305. doi: 10.1126/SCIENCE.287.5451.303

Zhang, J., Hansen, L. G., Gudich, O., Viehrig, K., Lassen, L. M. M., Schrübbers, L., et al. (2022). A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* 609, 341–347. doi: 10.1038/S41586-022-05157-3

Zhou, F., and Pichersky, E. (2020). More is better: the diversity of terpene metabolism in plants. *Curr. Opin. Plant Biol.* 55, 1–10. doi: 10.1016/J.PBI.2020.01.005

Zhu, X., Liu, X., Liu, T., Wang, Y., Ahmed, N., Li, Z., et al. (2021). Synthetic biology of plant natural products: from pathway elucidation to engineered biosynthesis in plant cells. *Plant Commun.* 2, 100229. doi: 10.1016/J.XPLC.2021.100229

Zorrilla-López, U., Masip, G., Arjó, G., Bai, C., Banakar, R., Bassie, L., et al. (2013). Engineering metabolic pathways in plants by multigene transformation. *Int. J. Dev. Biol.* 57, 565–576. doi: 10.1387/ijdb.130162pc

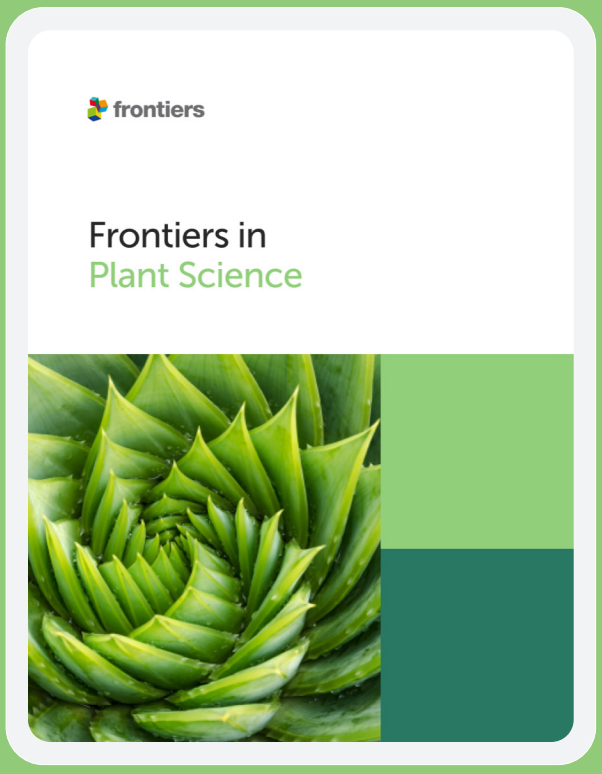
Frontiers in
Plant Science

Cultivates the science of plant biology and its applications

The most cited plant science journal, which advances our understanding of plant biology for sustainable food security, functional ecosystems and human health.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact