In Vivo Platforms for Terpenoid Overproduction and the Generation of Chemical Diversity

Guangkai Bian*, Tian Ma*, Tiangang Liu*,†,1

*Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education and School of Pharmaceutical Sciences, Wuhan University, Wuhan, PR China
†Hubei Engineering Laboratory for Synthetic Microbiology, Wuhan Institute of Biotechnology, Wuhan, PR China
[1]Corresponding author: e-mail address: liutg@whu.edu.cn

Contents

Abstract

Terpenoids represent a highly diverse group of natural products with wide applications. Engineering approaches have been used to increase titers of many value-added terpenoids, such as farnesene, taxadiene, lycopene, and astaxanthin. In this chapter, we review the in vitro reconstitution-based targeted engineering of terpenoids, as well as approaches for the mining of terpene cyclases and for increasing the chemical diversity. Information gained from in vitro reconstitution extends our understanding of the mechanisms underlying terpenoid biosynthesis, the contributions of enzymes and cofactors, and key enzymes and rate-limiting steps for the development of an ideal biosynthetic production system. The in vitro reconstitution-based targeted engineering strategy provides a rational and accurate engineering approach for terpenoid

Methods in Enzymology
ISSN 0076-6879
https://doi.org/10.1016/bs.mie.2018.04.025

overproduction with high efficiency. Furthermore, an efficient terpenoid overproduction platform can accelerate the entire process for the mining of terpene cyclases and the discovery of novel terpenoids and can substantially increase the chemical diversity of these kinds of terpenoids.

1. INTRODUCTION

Terpenoids are widely distributed in plants, microorganisms, insects, and various marine organisms. They represent the most diverse group of natural products. It has been estimated that more than 80,000 terpenoids have been structurally characterized (Pemberton et al., 2017).

Two distinct pathways, the methylerythritol phosphate (MEP) pathway and the mevalonate (MVA) pathway, in nature can produce the two universal 5-carbon isoprenoid building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), for the production of terpenoids. Subsequently, prenyltransferase assembles DMAPP and IPP to generate polyisoprenoid pyrophosphates with different lengths (C_{10}–C_{25}), such as geranyl diphosphate (GPP), farnesyl diphosphate (FPP), geranylgeranyl diphosphate (GGPP), and geranylfarnesyl diphosphate (GFPP). Terpenoids are then produced via terpene synthases using the polyisoprenoid pyrophosphates as substrates, resulting in various terpene skeletons (mono-, sesqui-, di-, sester-, tri-, tetra-, and polyterpenes). Finally, terpene skeletons are modified by tailor enzymes, such as cytochrome P450s, glycosyltransferase, and halogenase, to generate terpenoids with high structural diversity and complexity. Owing to their structural diversity, terpenoids have a wide range of applications; for example, they are used in food products, as antitumor and antimalaria agents, and as new jet fuel (Paddon et al., 2013; Wani, Taylor, Wall, Coggon, & McPhail, 1971; Zhu et al., 2015, 2014). However, the practical application of these value-added terpenoids is significantly hindered by low productivity and high cost.

Recent developments in metabolic engineering and synthetic biology provide alternative methods for producing value-added terpenoids in *Escherichia coli*, *Saccharomyces cerevisiae*, and filamentous fungi, resulting in several remarkable breakthroughs. The efficient overproduction of terpenoids, such as taxadiene (Ajikumar et al., 2010) and artemisinic acid (Paddon et al., 2013), has been realized by optimizing the precursor supply, balancing upstream and downstream pathways, balancing redox reactions, and introducing a dynamic regulatory system. To improve the convenience

and efficiency of engineering, the full reconstitution of the MVA pathway has been investigated using a systematic kinetic analysis. An in vitro reconstitution-based targeted metabolic engineering approach was developed for rationally and accurately engineering terpenoid overproduction with high efficiency (Zhu et al., 2014). Directed by this strategy, based on the well-engineered MVA pathway, high titers of farnesene, lycopene, taxadiene, and astaxanthin have been obtained using *E. coli* (Fig. 1), demonstrating the high efficiency and extensive applicability of the targeted engineering strategy (Bian, Yuan, Tao, Shi, & Liu, 2017; Ma et al., 2016; Zhu et al., 2015, 2014).

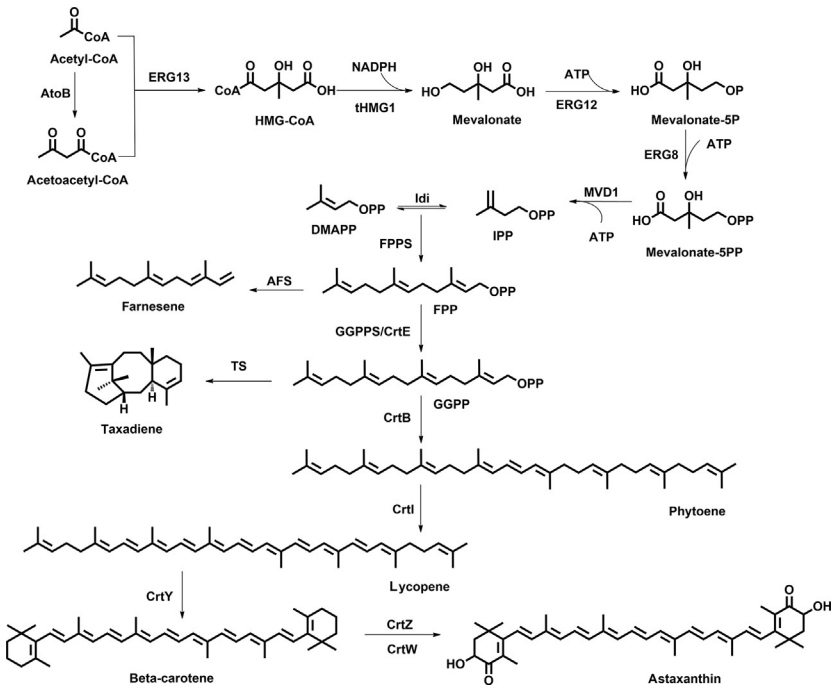

Fig. 1 Biosynthesis of farnesene, taxadiene, lycopene, and astaxanthin via the MVA pathway. *AFS*, α-farnesene synthase; *AtoB*, acetyl-CoA acetyltransferase; *CrtB*, phytoene synthase; *CrtI*, phytoene desaturase; *CrtW*, beta-carotene ketolase; CrtY, lycopene cyclase; *CrtZ*, beta-carotene hydroxylase; *DMAPP*, dimethylallyl pyrophosphate; *ERG8*, phosphomevalonate kinase; *ERG12*, mevalonate kinase; *ERG13*, HMG-CoA synthase; *FPP*, farnesyl pyrophosphate; *FPPS*, FPP synthase; *GGPP*, geranylgeranyl diphosphate; *GGPPS/CrtE*, GGPP synthase; HMG-CoA, (*S*)-3-hydroxy-3-methylglutaryl-CoA; *Idi*, IPP isomerase; *IPP*, isopentenyl diphosphate; *IspA*, FPPS from *E. coli*; *MVD1*, mevalonate pyrophosphate decarboxylase; *tHMG1*, the truncated 3-hydroxy-3-methylglutaryl-CoA reductase; *TS*, taxadiene synthase.

A large number of terpenoids have been structurally characterized, and it is unclear how many terpenoids exist in nature that are yet to be discovered. Various terpenoids as well as their biosynthetic elements have been characterized from diverse sources, such as the sesquiterpene synthase from *Streptomyces pristinaespiralis* (Klapschinski, Rabe, & Dickschat, 2016), the diterpene cyclases from *Hericium erinaceum* (Yang et al., 2017), and the sesterterpene synthase from plants and filamentous fungi (Bian, Han, Hou, Yuan, & Liu, 2017; Huang et al., 2017; Qin et al., 2015; Ye et al., 2015). The discovery of new terpene cyclases and new terpenoids provides a basis for understanding the biosynthetic mechanism, enriching the structural diversity, and screening for new bioactive compounds. An efficient terpenoid overproduction platform can accelerate the entire process for the mining of terpene cyclases and the discovery of terpenoids. Based on a well-developed terpenoid overproduction platform, the functions of promiscuous terpene cyclases FgMS (*Fusarium graminearum* mangicdiene synthase) and FgGS (*F. graminearum* GJ1012 synthase) were fully exploited to produce seven novel diterpenes and sesterterpenes with three unusual skeletons (Bian, Han, et al., 2017). Furthermore, by using well-engineered precursor production platforms, combined with a combinatorial biosynthesis strategy, a large number of terpenoids have been generated after the combination of class I and class II diterpene synthases from various sources (Andersen-Ranberg et al., 2016; Jia, Potter, & Peters, 2016). Consequently, a metabolic engineering-based terpenoid discovery strategy may accelerate the process of novel terpene skeletons identification and provide bioactive or high value-added compounds for drug discovery and industrial applications. In this chapter, we outline a targeted engineering strategy for the overproduction of terpenoids and the mining of terpene cyclases to increase chemical diversity.

2. IN VITRO RECONSTITUTION OF THE MVA PATHWAY AND TARGETED ENGINEERING OF VALUE-ADDED TERPENOIDS

2.1 In Vitro Reconstitution of the MVA Pathway

In vitro reconstitution is a powerful method to monitor the detailed information regarding enzymatic reaction mechanisms, steady-state kinetics, biochemical parameters, and the accumulation of intermediates (Lowry, Walsh, & Khosla, 2015). The information obtained by this method can

improve our understanding of the mechanism underlying terpenoid synthesis, the contributions of particular enzymes and cofactors, and the key enzymes and rate-limiting steps for an ideal biosynthetic production system. Directed by this strategy, researchers have realized the efficient overproduction of fatty acids and terpenoids efficiently in vitro and in vivo (Korman, Opgenorth, & Bowie, 2017; Liu et al., 2017; Yu, Liu, Zhu, & Khosla, 2011; Zhu et al., 2014).

For the in vitro reconstitution of the MVA pathway (Fig. 2), farnesene was selected as an indicator to evaluate the efficiency of genes related to this pathway. The *atoB*, *idi*, and *ispA* genes of *E. coli*, the *erg13*, *thmg1* (Polakowski, Stahl, & Lang, 1998), *erg12*, *erg8*, and *mvd1* (also known as *erg19*) of *S. cerevisiae*, and codon-optimized *afs* from *Malus* × *domestica* (Pechous & Whitaker, 2004) were selected and cloned into pET28a(+) for protein overexpression and purification (Zhu et al., 2014). Then, the plasmids for protein expression were individually transformed into *E. coli* BL21 (DE3). The proteins were purified and maintained in storage buffer (100 m*M* phosphate buffer, 10% glycerol, pH 7.6).

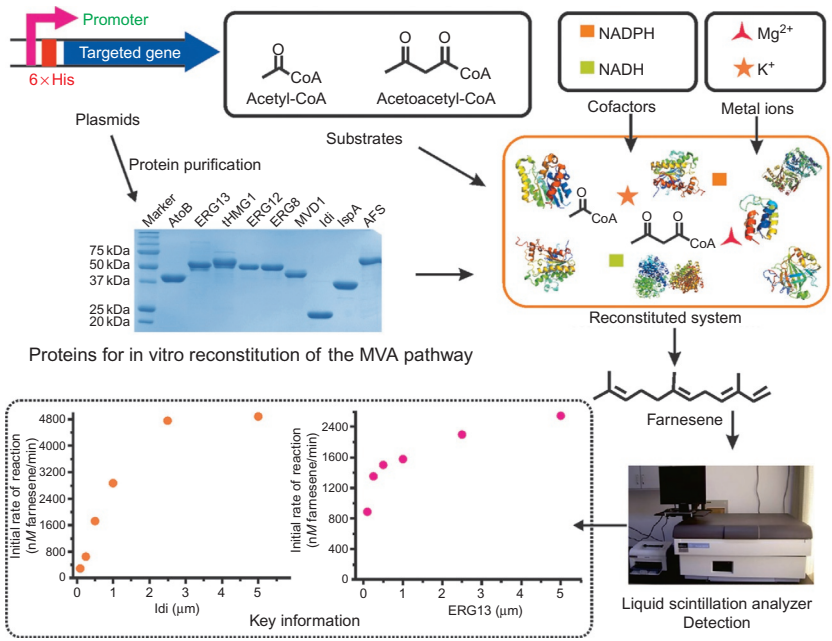

Fig. 2 In vitro reconstitution of MVA pathway.

An in vitro reconstitution assay was performed and the production of farnesene was detected by GC/MS and radioactive thin-layer chromatography. Based on a kinetic analysis, increasing the concentrations of ERG13, ERG8, IspA, and AFS from 0.1 to 5 μ*M* can mildly (1.5- to 2.7-fold) promote the production of farnesene. Increasing the concentration of Idi (0.1–5 μ*M*) can substantially increase the initial rate of farnesene production by 17-fold (Fig. 2), indicating the key role of Idi in terpenoid engineering. Additionally, an analysis of synergistic effects showed that simultaneously increasing the concentrations of Idi together with ERG13, ERG8, and AFS can increase the production of farnesene by fivefold. Based on the results of titration studies, the molar ratio of AtoB:ERG13:tHMG1:ERG12:ERG8:MVD1:Idi:IspA:AFS was adjusted slightly to 1:10:2:5:5:2:5:2:2 (Zhu et al., 2014). These previous studies provide an ultimate goal for metabolic engineering to realize the overproduction of terpenoids with high efficiency.

Instead of constructing a series of mutants to evaluate the contributions of enzymes to terpenoid yield, key information from in vitro assays provides direction for the engineering of terpenoid production in microorganisms. Thus, this approach provides a fantastic strategy for the realization of the overproduction of value-added natural products in a convenient and efficient manner.

2.1.1 Protein Expression and Purification for In Vitro Reconstitution

The procedures for protein expression and purification for in vitro reconstitution can be summarized as follows.

1. Plasmids related to the expression of each enzyme in the MVA pathway and the downstream farnesene production pathway are transformed into *E. coli* BL21 (DE3) for protein purification.
2. Mutants are cultivated in 2-L flasks containing 1 L of LB medium at 37°C with 50 mg/L kanamycin (Kan).
3. When the OD_{600} reaches 0.6–0.8, 0.1 m*M* IPTG is added to the culture, followed by cultivation for an additional 18 h at 16°C.
4. Cells are harvested and resuspended in 40 mL of buffer A (50 m*M* Tris–Cl, 300 m*M* NaCl, 4 m*M* β-mercaptoethanol, pH 7.6).
5. A high-pressure homogenizer is used to lyse the cell suspensions, followed by centrifugation at 30,000 × *g* for 1 h.

6. The supernatant is filtered and supplemented with 6% buffer B (500 m*M* imidazole in buffer A) to adjust the supernatant to 30 m*M* imidazole to reduce nonspecific binding.
7. The samples are subjected to metal affinity chromatography using a Ni-NTA column.
8. The sample is supplemented with 8% buffer B to rinse unbound and nonspecifically bound proteins.
9. Buffer B is linearly increased to 60% in 12 CVs (column volumes) and then to 100% for an additional 4 CVs.
10. SDS-PAGE is used to detect the targeted proteins and their purity.
11. Amicon Ultra-15 centrifugal filter devices are used to concentrate the purified target protein to a final volume of 2.5 mL.
12. The PD-10 column is equilibrated with storage buffer (PB buffer: 100 m*M* phosphate buffer, 10% glycerol, pH 7.4), and the concentrated protein is loaded to change buffer.
13. A Pierce BCA Protein Assay Kit and EnSpire Multimode Plate Reader are used to measure the protein concentration.
14. Finally, proteins are stored at −80°C after flash freezing in liquid nitrogen.

2.1.2 Extraction and Detection of Farnesene

To detect the production of farnesene, the reaction mixture was extracted by the following procedures:

1. Extract the reaction mixture with an equal volume of hexane.
2. Centrifuge the hexane layer at 10,000 × *g* for 5 min and then collect the organic layer for GC/MS analysis.
3. Inject the samples into a TRACE TR-5MS volume (30 m × 0.25 mm × 0.25 μm). Set the oven temperature at 80°C for 1 min, increased to 220°C at a rate of 10°C/min, and held at 220°C for 15 min. Maintain the injector and transfer lines at 230°C and 240°C, respectively.

2.2 Targeted Engineering of Farnesene

Farnesene is one of the most important sesquiterpenes with wide applications, e.g., as a precursor of next-generation jet fuel, lubricants, solvents, polymers, and vitamin E (https://amyris.com/products/polymers/).

For the targeted engineering of farnesene, different from the traditional metabolic engineering strategy that requires the construction of a large number of mutants, information from the in vitro reconstitution system can indicate the rate-limiting steps in the whole pathway and can be used to establish target yields. Directed by this information, and together with targeted proteomic analyses and mass spectrometry-based intermediate analyses, inefficiencies in the concentrations of ERG13, ERG8, Idi as well as the accumulation of ATP, NADPH, and the intermediates Acetyl-CoA, MVA, IPP/DMAPP, and FPP were clearly identified, providing a basis for further metabolic engineering research (Zhu et al., 2014). Based on these findings, the expression levels of ERG13, ERG8, and Idi were increased. As a result, the production of farnesene increased predictably stepwise to 1.1 g/L in strain F4 (2000-fold) at the shake-flask scale (Fig. 3) (Zhu et al., 2014).

2.2.1 Plasmids for the MVA Pathway and Farnesene Overexpression

To realize the overproduction of the MVA pathway, the plasmids pMH1 and pFZ81 with genes in the MVA pathway and plasmid pFZ71 with the downstream farnesene production pathway were constructed by Simple Cloning (You, Zhang, & Zhang, 2012) and Gibson method (Gibson et al., 2009). A detailed description of the construction procedure is as follows.

For plasmid pMH1 construction, the origin of pBBR1MCS-1 was replaced with p15A before the insertion of the indicated genes. The primers are listed in Table 1. The p15A origin was amplified from pMSD15 using the primer pair P1/P2, and pBBR1MCS1 was simultaneously amplified using the primer pair P3/P4. The two PCR fragments were combined using the Simple Cloning method to obtain the plasmid pBBR1MCS1/p15A. Gibson method was then used to assemble the indicated genes into pBBR1MCS1/p15A to yield pMH1. First, the three primer pairs P5/P6, P7/P8, and P9/P10 were used to amplify *atoB*, *erg13*, and *thmg1*, respectively, and the primers P11/P12 were used to amplify the pBBR1MCS1/p15A plasmid backbone. Finally, the pBBR1MCS1/p15A PCR fragment, *atoB*, *erg13*, and *thmg1* were assembled by the Gibson method to generate pMH1. For plasmid pFZ81 construction, *erg12*, *erg8*, *mvd1*, and *idi* were amplified from the expression vector using the primers P13/P14, P15/P16, P17/P18, and P19/P20, respectively. pBBR1MCS2 was amplified

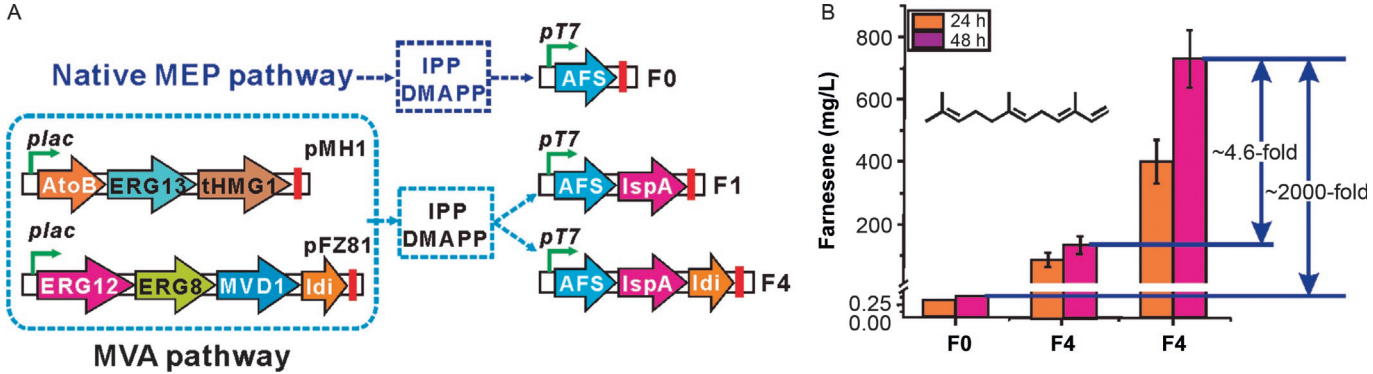

Fig. 3 Overproduction of farnesene in *E. coli* (Zhu et al., 2014). (A) Production of α-farnesene by the endogenous MEP pathway or synthetic MVA pathways and depiction of the synthetic operons used in this study. (B) Production of farnesene in engineered strains.

Table 1 Primers for In Vitro Reconstitution of the Mevalonate Pathway, Terpenoid Overproduction, and Functional Characterization of Terpene Cyclases

Primers	Sequence 5′–3′
P1	GCATTGCCGCCGTGGGTTTCTCGAGCGGGGCGGAGATTTCCTGGAAGATG
P2	GGGGCCACTTTTTGCCGGAGCTCGAGAAATATTTTATCTGATTAATAAGATG
P3	CATCTTATTAATCAGATAAAATATTTCTCGAGCTCCGGCAAAAAGTGGCCCC
P4	CATCTTCCAGGAAATCTCCGCCCCGCTCGAGAAACCCACGGCGGCAATGC
P5	CCCGGGGATCCTCTAGAGTCGACTAGGAGGAATATAAAATGAAAAATTGTGTCATCGTC
P6	GTTGAGAGTTTCATTTAGCTGTCCTCCTTAATTCAACCGTTCAATCACC
P7	ACGGTTGAATTAAGGAGGACAGCTAAATGAAACTCTCAACTAAACT
P8	TGGCTGCTGCCCATAGTGTAATCCTCCTTATTTTTTAACATCGTAAGAT
P9	ATGTTAAAAAATAAGGAGGATTACACTATGGGCAGCAGCCATCATCA
P10	TTAGGATTTAATGCAGGTGACGG
P11	TTTGAAAGATGGGTCCGTCACCTGCATTAAATCCTAAGGATCCACTAGTTCTAGA
P12	TTTTATATTCCTCCTAGTCGACTCTAGAGGATCCCCGGGCTGCAGGAATTCGATATCAA
P13	CTGCAGTAGGAGGAATTAACCATGTCATTACCGTTCTTAACT
P14	CTCAACTCTGACATTTGATCTGCCTCCTATGAAGTCCATGGTAA
P15	TTACCATGGACTTCATAGGAGGCAGATCAAATGTCAGAGTTGAGAGCCTTC
P16	GATGCTGTGTAAACGGTCATGAGTATTACCTCCTATTTATCAAGATAAGTTTC
P17	ATCTTGATAAATAGGAGGTAATACTCATGACCGTTTACACAGCATCCGTT
P18	TGCCCATATAGTAATCCTCCTCCCGGGCTGCAGTTATTCCTTTGGTAGACCAGTC
P19	GGAATAACTGCAGCCCGGGAGGAGGATTACTATATGGGCAGCAGCCATCAT
P20	TTATTTAAGCTGGGTAAATGCAGA
P21	CAGAAAACGATTATCTGCATTTACCCAGCTTAAATAAGAGCTCCAATTCGCCCTATAGT
P22	TTAAGAACGGTAATGACATGGTTAATTCCTCCTACTGCAGGAATTCGATATCAAG
P23	ACCCAGCTTAAATAAAGGAGGATTACACTATGCCGCG GTATGATCTGATTC
P24	GCATTCCAAATCCACAACATATAGTAATCCTCCTTCAT TGCATCGCCTGTTGAC
P25	CAGGCGATGCAATGAAGGAGGATTACTATATGTTGTG GATTTGGAATGCCCTGA
P26	CCACTGCGGCGGTCATTACTCATTCCTCCTTTACTTCC CGGGTGGCGCGTC
P27	CGCCACCCGGGAAGTAAAGGAGGAATGAGTAATGAC CGCCGCAGTGGCAGAG
P28	GCGGTTTCTTTACCAGACTCGAGTTAGCTCTCACCACG CCATAAG
P29	GAAATAATTTTGTTTAACTTTAAGAAGGAGATATACC ATGACCGTGTGTGCGAAAAAAC
P30	TCACCGTGGTCGGTTTCATGGTTAATTCCTCCTTTACG ACACCGCTGCCAG
P31	CTGGCAGCGGTGTCGTAAAGGAGGAATTAACCATGAA ACCGACCACGGTGA
P32	TCAGCAGGGACGGATTATTCATGAGTATTACCTCCTTT AAATCAGGTCTTCCAGCATC
P33	GATGCTGGAAGACCTGATTTAAAGGAGGTAATACTCATGAATAATCCGTCCCTGCTGA

Table 1 Primers for In Vitro Reconstitution of the Mevalonate Pathway, Terpenoid Overproduction, and Functional Characterization of Terpene Cyclases—cont'd

Primers	Sequence 5′–3′
P34	TTGTCGACCTGCAGGCGCGCCGAGCTCGAATTCACTA GTTAAACGGGGCGCTGCCAGAG
P35	TTGCGTTTCGTCAGAACTGAGTCGACAAGCTTGCGGCCG
P36	GTACCCGTAGAGCTAGACATGGTATATCTCCTTCTTAAAG
P37	CTTTAAGAAGGAGATATACCATGTCTAGCTCTACGGGTAC
P38	ACATATGTATATCTCCTTCTTTAGACCTGGATTGGATCGAT
P39	ATCGATCCAATCCAGGTCTAAAGAAGGAGATATACATATGT
P40	CGGCCGCAAGCTTGTCGACTCAGTTCTGACGAAACGCAA
P41	ATATCATATGGATTTCACCTACCGTTATAG
P42	ATATCTCGAGACTAGTTAGTCGATACGCAGCATCTC
P43	GCACGAGCTCATGGATCCCTACAGTGAAACATCAG
P44	AGTCAAGCTTACTAGTCAACCCTCGTGCTGACAAGC
P45	ATATGGATCCATGGACTTTCCGCAGCAACTC
P46	ATATGAATTCACTAGTTTATTTATTACGCTGGATGATG
P47	ATCGCATATGCAAACGGAACACGTCATTTTATTG
P48	ATATCTCGAGACTAGTTATTTAAGCTGGGTAAATGC
P49	CTTTAAGAAGGAGATATACCATGGATTTCACCTACCGTTATAG
P50	GCCATATGTATATCTCCTTCTTTAGTCGATACGCAGCATCTC
P51	GAGATGCTGCGTATCGACTAAAGAAGGAGATATACATATGGCGCAGCTGAGCG
P52	CTGTTCGACTTAAGCATTATGCGGCCGCAAGCTTGTCGACTTAGTGGTCACGCGCCGCAACC
P53	GTCGACAAGCTTGCGGCCGCATAATGCTTAAGTCGAACAG
P54	ATAACGGTAGGTGAAATCCATGGTATATCTCCTTCTTAAAG
P55	CTTTAAGAAGGAGATATACCATGGATCCCTACAGTGAAAC
P56	CATATGTATATCTCCTTCTTCAACCCTCGTGCTGACAAGCC
P57	GCTTGTCAGCACGAGGGTTGAAGAAGGAGATATACATATGTTTGATTTCAATGAATATATG
P58	CTGTTCGACTTAAGCATTATGCGGCCGCAAGCTTGTCGACTCAGTTCTGACGAAACGCAATG
P59	GTTTCACTGTAGGGATCCATGGTATATCTCCTTCTTAAAG

by PCR using the primers P21/P22. Finally, *erg12*, *erg8*, *mvd1*, *idi*, and pBBR1MCS2 were assembled by the Gibson method to obtain pFZ81. For farnesene overproduction, the *Xba*I–*Xho*I fragment of *afs* from pFZ31 was inserted into pET21a(+) to generate pFZ35, and the *Xba*I–*Xho*I fragment of *ispA* from pFZ32 was ligated into the *Spe*I–*Xho*I fragment of pFZ35 to generate pFZ38. An additional copy of *idi* was introduced by inserting the *Xba*I–*Xho*I fragment of *idi* into the *Spe*I–*Xho*I fragment of pFZ38 to generate pFZ71 under the control of the T7 promoter. Plasmids are listed in Table 2.

Table 2 Strains and Plasmids Used for the In Vitro Reconstitution of the Mevalonate Pathway, Terpenoid Overproduction, and Functional Characterization of Terpene Cyclases

Strains	Relevant Genotype	References
BL21 (DE3)	*E. coli B F⁻ dcm ompT hsdSB(rB⁻ mB⁻) gal*	Invitrogen
K12 MG1655 (DE3)	F⁻ λ⁻ ilvG-rfb-50 rph-1 (DE3)	Invitrogen
F4	*E. coli* BL21:: pMH1, pFZ81, pFZ71	Zhu et al. (2014)
L3	*E. coli* BL21:: pMH1, pFZ81, pFZ111	Zhu et al. (2015)
A1	*E. coli* K12 MG1655: pMH1, pFZ81, pFZ153	Ma et al. (2016)
T4	*E. coli* BL21:: pMH1, pFZ81, pFZ131	Bian, Yuan, et al. (2017)
T7	*E. coli* BL21:: pMH1, pFZ81, pGB309	Bian, Han, et al. (2017)
T11	*E. coli* BL21:: pMH1, pFZ81, pGB313	Bian, Han, et al. (2017)
T12	*E. coli* BL21:: pMH1, pFZ81, pGB314	Bian, Han, et al. (2017)

Plasmids	Origin	Description	References
pGZI	pBR322	pT7: N-terminal His-tagged Idi, Kan^+	Zhu et al. (2014)
pJW1	pBR322	pT7: N-terminal His-tagged AtoB, Kan^+	Zhu et al. (2014)
pFZ12	pBR322	pT7: N-terminal His-tagged ERG13, Kan^+	Zhu et al. (2014)
pJW2	pBR322	pT7: N-terminal His-tagged tHMG1, Kan^+	Zhu et al. (2014)
pFZ13	pBR322	pT7: N-terminal His-tagged ERG12, Kan^+	Zhu et al. (2014)
pFZ14	pBR322	pT7: N-terminal His-tagged ERG8, Kan^+	Zhu et al. (2014)
pFZ15	pBR322	pT7: N-terminal His-tagged MVD1, Kan^+	Zhu et al. (2014)

Table 2 Strains and Plasmids Used for the In Vitro Reconstitution of the Mevalonate Pathway, Terpenoid Overproduction, and Functional Characterization of Terpene Cyclases—cont'd

Plasmids	Origin	Description	References
pFZ31	pBR322	pT7: N-terminal His-tagged AFS, Kan^+	Zhu et al. (2014)
pFZ32	pBR322	pT7: N-terminal His-tagged IspA, Kan^+	Zhu et al. (2014)
pFZ35	pBR322	pT7: AFS, Amp^+	Zhu et al. (2014)
pFZ36	pBR322	pT7: N-terminal His-tagged IspA, Kan^+	Zhu et al. (2014)
pFZ38	pBR322	pT7: AFS, N-terminal His-tagged IspA, Amp^+	Zhu et al. (2014)
pFZ71	pBR322	pT7: AFS, N-terminal His-tagged IspA and Idi, Amp^+	Zhu et al. (2014)
pFZ81	pBBR1MCS	plac: ERG12, ERG8, MVD1 and N-terminal His-tagged Idi, Kan^+	Zhu et al. (2014)
pMH1	p15A	plac: AtoB, ERG13 and N-terminal His-tagged tHMG1, Chl^+	Zhu et al. (2014)
pFZ21	pBR322	pT7: N-terminal His-tagged CrtE, Kan^+	Zhu et al. (2015)
pFZ22	pBR322	pT7: N-terminal His-tagged CrtB, Kan^+	Zhu et al. (2015)
pFZ23	pBR322	pT7: N-terminal His-tagged CrtI, Kan^+	Zhu et al. (2015)
pFZ27	pBR322	pT7: N-terminal His-tagged CrtE, CrtB and CrtI, Kan^+	Zhu et al. (2015)
pFZ110	pBR322	pT7: N-terminal His-tagged CrtE, CrtB and CrtI, Amp^+	Zhu et al. (2015)
pFZ111	pBR322	pT7: N-terminal His-tagged CrtE, CrtB, CrtI and Idi, Amp^+	Ma et al. (2016)
pFZ112	pBR322	pT7: CrtE, CrtI, CrtB, Amp^+	Ma et al. (2016)
pFZ153	pBR322	pT7: CrtE, CrtI, CrtB, pT7: Idi, CrtY, CrtZ, CrtW, Amp^+	Ma et al. (2016)

Continued

Table 2 Strains and Plasmids Used for the In Vitro Reconstitution of the Mevalonate Pathway, Terpenoid Overproduction, and Functional Characterization of Terpene Cyclases—cont'd

Plasmids	Origin	Description	References
pFZ131	pBR322	pT7: TS, GGPPS, Amp^+	Bian, Yuan, et al. (2017)
pMH1	p15A	plac: AtoB, ERG13, tHMG1, Chl^+	Zhu et al. (2014)
pFZ81	pBBR1MCS	plac: ERG12, ERG8, MVD1, Idi, Kan^+	Zhu et al. (2014)
pGB309	pBR322	pT7: FgMS, GFPPS; pT7: Idi, Amp^+	Bian, Han, et al. (2017)
pGB313	pBR322	pT7: FgGS, GGPPS; pT7: Idi, Amp^+	Bian, Han, et al. (2017)
pGB314	pBR322	pT7: FgGS, FPPS, Idi, Amp^+	Bian, Han, et al. (2017)

2.2.2 Shake-Flask Fermentation and Analysis of α-Farnesene Production in Engineered Strains

To assess farnesene production in shake flasks, the engineered *E. coli* F4 (Table 2) was cultured in 500-mL flasks containing 150 mL of 2× TY medium at 30°C with 2% glycerol, 100 mg/mL ampicillin (Amp), 50 mg/mL Kan, and 34 mg/mL chloramphenicol (Chl). When the OD_{600} reached 0.6–0.8, IPTG was added to a final concentration of 0.1 m*M*. After 3 h of induction, the medium was overlaid with 30 mL of decane. Samples containing 1 mL of decane and 5 mL of culture were withdrawn at designated time points. After centrifugation, the decane phase was subjected to GC/MS according to the method described in Section 2.1.2. The production α-farnesene was quantified by GC/MS; trans-β-farnesene in decane was used as a calibration standard as described previously (Wang et al., 2011).

2.3 Targeted Engineering of Lycopene

Lycopene is a well-known tetraterpenoid with attractive antioxidative, anticancer, and antiinflammatory activities; it has been used as a pharmaceutical compound, nutraceutical, functional food, and cosmetic additive

(Zhu et al., 2015). Extensive efforts have focused on increasing the production efficiency of lycopene in *E. coli* and *S. cerevisiae* (Coussement, Bauwens, Maertens, & De Mey, 2016; Sun et al., 2014; Xie, Lv, Ye, Zhou, & Yu, 2015). The biosynthesis of lycopene begins with the condensation of IPP and DMAPP, which are catalyzed by CrtE to form GGPP, and then GGPP is catalyzed by CrtB to produce phytoene, and phytoene is catalyzed by CrtI to produce lycopene (Fig. 4).

With a clear biochemical basis and few rate-limiting steps, the MVA pathway is widely used for carotenoid overproduction in *E. coli*. The above-mentioned targeted engineering strategy has been employed to accelerate the engineering process and efficiently produce lycopene. By taking advantage of the efficient precursor-producing platform (*E. coli* BL21 (DE3)/pMH1/pFZ81) constructed for farnesene production together with the downstream lycopene-forming pathway (pFZ111), the titer of lycopene in strain *E. coli* L3 was increased to 34.3 mg/g DCW (dry cell weight) (1.23 g/L) in a 100-L fermenter (Zhu et al., 2015). This indicates that the information obtained from the targeted engineering strategy is universal and can be utilized to realize the overproduction of other terpenoids with limited effort in a short period of time.

2.3.1 Plasmids for Lycopene Overproduction

To produce lycopene in *E. coli*, *crtE*, *crtB*, and *crtI* from *Pantoea ananatis* (Misawa et al., 1990) were codon-optimized, synthesized, and inserted into pET28a(+) to produce pFZ21, pFZ22, and pFZ23, respectively. Subsequently, the *Xba*I–*Xho*I fragments of *crtB* and *crtI* were sequentially inserted into pFZ21 to produce pFZ27. The *Xba*I–*Xho*I fragment of pFZ27 was inserted into pET21a(+) to produce pFZ110. An additional copy of *idi* was introduced into pFZ110 by inserting the *Xba*I–*Xho*I fragment of *idi* into the *Spe*I–*Xho*I fragment of pFZ110 to produce pFZ111.

2.3.2 Analysis of Lycopene Titers in Engineered Strains

Overproduction of lycopene is a fascinating work, and many researchers have realized the high efficiency production of lycopene in *E. coli*, for example, Zhu et al. reporting a concentration of 34.3 mg/g DCW (Zhu et al., 2015) and Sun et al. reporting 50.6 mg/g DCW (Sun et al., 2014). However, different quantitative assessment method used can result in significant differences in the yield of lycopene and may mislead the metabolic engineering process. A closer examination of the study by Coussement

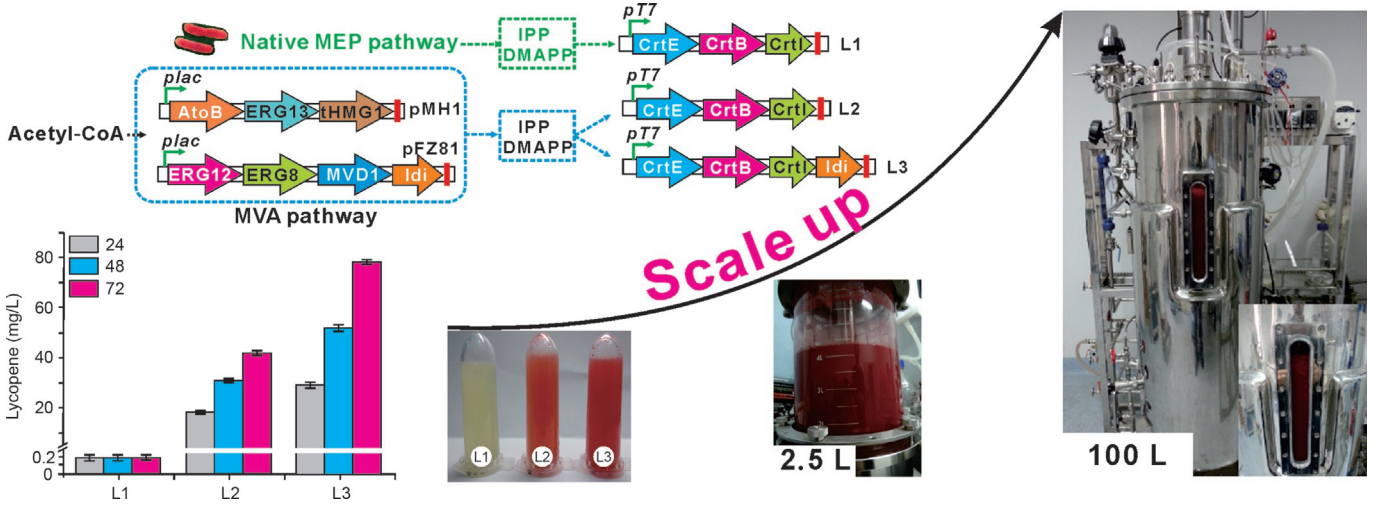

Fig. 4 Targeted engineering of lycopene overproduction in *E. coli*. Production of lycopene via endogenous MEP pathway or synthetic MVA pathway and depiction of the synthetic operons used in this study. Production of lycopene in engineered strains L1–L3 at shake-flask level was detected, and the fed-batch fermentation of strain L3 was scaled up to a 100-L fermenter.

et al. (2016), which yielded 448 mg/g DCW, revealed several notable details: the reference strain should not produce any lycopene, but it already showed a high production level; after incubating the 96-well plate containing the extraction solution at 50°C for 10 min, the total solvent decreased to <100μL, and although this volume would not meet the requirements for the next operation, there were no details on how to avoid reagent volatilization. Most importantly, though, no detailed genetic information was provided regarding the highest production strain, which has prevented other groups from repeating this experiment and building on the research. The significant differences in quantitative assessment of lycopene production, raises questions regarding how assessments are made and the best way to compare different strategies.

The extraction and titration method we recommend is as follows:

1. Harvest 1 mL cell culture of engineered strain and centrifuge at $5000 \times g$ for 5 min, remove the supernatants completely.
2. Vortex to disperse the cell pellet, then use acetone (with 1% butylated hydroxytoluene, BHT) to extract the pellets at 55°C for 15 min in the dark, with vortexing for 1 min every 5 min.
3. Centrifuge the solution at $13{,}000 \times g$ for 10 min.
4. Transfer the supernatant to 96-well polypropylene plates for direct UV absorbance measurements or to sample vials for HPLC measurements.
5. Calculate the titer of lycopene using a standard curve with appropriate dilution factor and averaged for three replicate analyses as described previously (Alper, Jin, Moxley, & Stephanopoulos, 2005).
6. Count the cell mass by OD_{600} with a coefficient of 0.3 g DCW/OD_{600}.

For lycopene standard calibration, the content of lycopene in the standard was corrected with an extinction coefficient of 3450 ($A^{1\%} = 3450$) at 472 nm (hexane as solvent), as described previously (Joint FAO/WHO Expert Committee on Food Additives, 2006). The details are as follows: ~5 mg of the standard was weighed in a 100-mL volumetric flask, dissolved in 5 mL of tetrahydrofuran (0.025% BHT), and then supplement with hexane to a final volume of 100 mL. Next, 5.0 mL of the solution was transferred into a 100-mL volumetric flask, and then supplement with hexane to a final volume of 100 mL, producing the standard solution for UV detection. The absorbances of the standard and candidate solutions were measured in 1-cm cuvettes at the wavelength of maximum absorption (approximately 472 nm). Hexane was used as the blank.

$$\text{Calculations}: C_{St}\,(\text{mg/L}) = A_{472} \times 10{,}000/3450$$

where C_{St} is the lycopene concentration, A_{472} is the absorbance at the wavelength of lycopene maximum absorption in hexane, and 3450 is the specific absorbance $A^{1\%}$ of lycopene in hexane. The scaling factor is 10,000.

$$\text{Lycopene } \% = C_{St} \times 20 \times 0.1/W$$

where lycopene % is the percentage of lycopene in the standard, 20 is the dilution ratio, 0.1 is the solvent volume, and W is the weight of the sample.

There are two methods for lycopene titer analyzing. For UV detection method, detection of lycopene can be carried out using the Perkin Elmer EnSpire Multimode plate reader at 474 nm (acetone as solvent) with 150 μL acetone extraction solution in polypropylene 96-well plates. For HPLC detection method, an HPLC system (Thermo Scientific, Dionex, UltiMate 3000) equipped with a C18 column (4.6 × 150 mm, 5 μm, Agilent Technologies, Inc.) was used to detect the 474 nm signals at 30°C. Samples were eluted with solvent A [90% aqueous acetonitrile (HPLC grade)] and solvent B [methyl alcohol-isopropyl alcohol (3:2, v/v, HPLC grade)] according to the following procedure with flow rate set as 1.0 mL/min: 0–15 min, 100%–10% solvent A, 0%–90% solvent B; 15–30 min, 10% solvent A, 90% solvent B; 30–35 min, 10%–100% solvent A, 90%–0% solvent B (Zhou, Ye, Xie, Lv, & Yu, 2015). Both detection methods are recommended; however, UV is more suitable for high-throughput lycopene detection and HPLC is preferred for quantification of lycopene content in mixture.

2.4 Targeted Engineering of Astaxanthin

Astaxanthin is a kind of cartenoid with various beneficial effects on human health; it has widely used in nutritional supplements, pharmaceuticals, cosmetics, food colorants, and animal feed additives (Fraser & Bramley, 2004; Tominaga, Hongo, Karato, & Yamashita, 2012). Astaxanthin can be produced by using the lycopene production pathway in addition to the downstream astaxanthin production pathway consisting of CrtY, CrtZ, and CrtW. CrtY can catalyze the transformation of lycopene to beta-carotene, and CrtZ together with CrtW can catalyze the formation of astaxanthin by using β-carotene as a substrate (Fig. 1). Consequently, similar to the above-mentioned overproduction of farnesene and lycopene, an efficient precursor production pathway plays a significant role in the overproduction of astaxanthin.

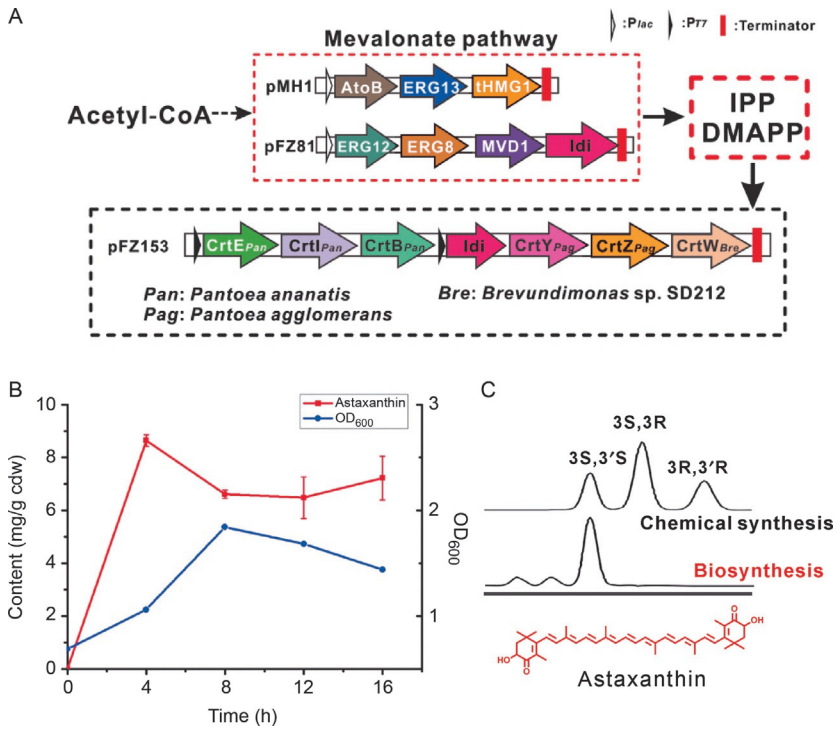

Fig. 5 Targeted engineering of astaxanthin overproduction in *E. coli*. (A) *E. coli* A1 with pMH1, pFZ81, and pFZ153 for astaxanthin overproduction. (B) Content of astaxanthin produced by *E. coli* A1. (C) Configuration of astaxanthin produced by chemical synthesis and biosynthesis in our system.

Using the targeted engineering strategy, an efficient astaxanthin overproduction mutant *E. coli* A1 (Table 2) was constructed by employing the same farnesene precursor-providing platform, together with *CrtEIB*, *CrtYZ*, *CrtW*, and an additional copy of *Idi* from *E. coli*. The specificity production of 3S, 3′S-astaxanthin was realized (3S, 3′S configuration has the highest biological activity), and the highest yield of 8.64 mg/g DCW was obtained in *E. coli* (Fig. 5) (Ma et al., 2016).

2.4.1 Plasmid Construction for Astaxanthin Overproduction

For astaxanthin overproduction, a plasmid harboring the downstream astaxanthin-forming pathway was constructed. Primers used for astaxanthin overproduction are listed in Table 1. The *idi* gene from pGZI (Zhu et al., 2014) was digested by *Nde*I and *Xho*I and cloned into the plasmid

pETDuet-1 to generate pFZ87. The *crtY* and *crtZ* genes were amplified individually from *Pantoea agglomerans* (Hundle et al., 1994; Schnurr, Schmidt, & Sandmann, 1991) using the primer pairs P23/P24 and P25/P26. The *crtW* gene from *Brevundimonas* sp. SD212 (Nishida et al., 2005) was codon optimized, synthesized, and amplified using the primer pair P27/P28. The four fragments containing *idi*, *crtY*, *crtZ*, and *crtW* were assembled by the Gibson method to obtain the plasmid pFZ152. The *crtE*, *crtI*, and *crtB* genes were separately amplified from pFZ21, pFZ22, and pFZ23 with the primer pairs P29/P30, P31/P32, and P33/P34, respectively (Zhu et al., 2015). These three fragments were assembled by the Gibson method to generate pFZ112. Finally, the *Nde*I–*Eco*RI digested fragment of pFZ112 (including *crtE*, *crtI*, and *crtB*) was inserted into pFZ152 to produce pFZ153. The constructed plasmids are listed in Table 2.

2.4.2 Fermentation and Quantification of Astaxanthin Produced by E. coli

E. coli K-12 MG1655 (DE3) was selected as a host for astaxanthin overproduction. During the shake-flask fermentation period, *E. coli* A1 (*E. coli* K12 MG1655/pMH1/pFZ81/pFZ153) was cultured and induced by IPTG as described in Section 2.2.2. The production of astaxanthin was achieved and detected according to the following procedure.

1. The cell culture (2 mL) is harvested by centrifugation at $10{,}000 \times g$ for 5 min.
2. The cell pellet is dispersed by vortexing, and methanol and acetone (1:4, v:v) are used to extract the pellets in the dark until the residues become colorless.
3. The supernatant is pooled, centrifuged, and transferred to sample vials for HPLC detection.
4. Samples are detected by the HPLC system according to the procedure described in Section 2.3.2.

2.5 Targeted Engineering of Taxadiene

As described earlier, the targeted engineering of terpenoids by the MVA pathway has been successfully conducted in *E. coli* for the production of the sesquiterpene product farnesene and the tetraterpene compounds lycopene and astaxanthin (Ma et al., 2016; Zhu et al., 2015, 2014). The efficiency of this platform for diterpene overproduction in *E. coli* was also evaluated. Taxadiene, a precursor of one of the most important anticancer drugs, contains a C20 backbone derived from the condensation and

cyclization of the universal C5 units IPP and DMAPP. To test the efficiency of a targeted engineering strategy for diterpene overproduction in *E. coli*, the downstream GGPPS, the taxadiene synthase, and an additional copy of Idi were introduced to generate the pFZ131 plasmid for taxadiene production (Fig. 6) (Bian, Yuan, et al., 2017).

2.5.1 Plasmid Construction for Taxadiene Overproduction

For plasmid pFZ131 construction, the backbone of the pETDuet-1 fragment was amplified using the primer pair P35/P36 (Table 1), and the *ts* and *ggpps* genes were synthesized as described by Ajikumar and amplified by P37/P38 and P39/P40, respectively (Ajikumar et al., 2010). The three fragments were assembled by the Gibson method to generate pFZ131 (Gibson et al., 2009).

2.5.2 Overproduction and Detection of Taxadiene in E. coli

For taxadiene overproduction in *E. coli*, strain *E. coli* T4 (Table 2) was constructed and cultivated in 500-mL shake flasks according to the procedure described in Section 2.2.2, with slight modifications, including cultivation at 28°C for an additional 3 days after induction with IPTG. Production of taxadiene was detected by GC–MS following the method described in Section 2.1.2. The cell concentration was measured by OD_{600} with a coefficient of 0.3 g DCW/OD_{600}. As a result, taxadiene was produced by *E. coli* T4, with a titer of 11.3 mg/L (13.2 g/g DCW) at the shaking flask level (Bian, Yuan, et al., 2017). This demonstrates the wide applicability of the targeted engineering strategy.

3. GENOME MINING OF TERPENE CYCLASES AND THE GENERATION OF CHEMICAL DIVERSITY

The development of genome sequencing and various bioinformatics tools, such as antiSMASH and SMURF (Khaldi et al., 2010; Weber et al., 2015), have enabled the characterization of a large number of terpene cyclases and have substantially increased the chemical diversity of terpenoids (Fig. 7) (Bian, Han, et al., 2017; Bian et al., 2018; Yamada et al., 2015). However, there are still many terpene cyclases were unknown functions. The well-established terpenoid overproduction platform makes it possible to efficiently characterize the functions of terpene cyclases, irrespective of the organism and type of terpenoid. This allows us to avoid the negative

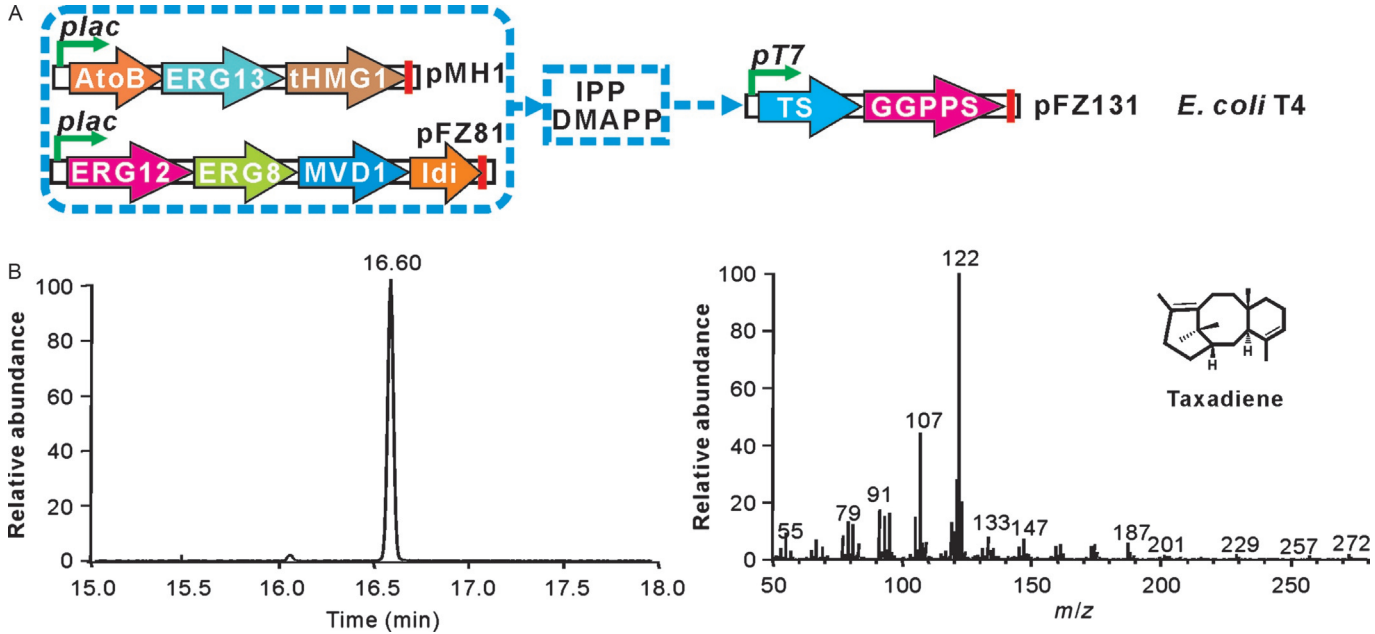

Fig. 6 Production of taxadiene in *E. coli*. (A) *E. coli* T4 for taxadiene overproduction. (B) Detection of taxadiene by GC–MS.

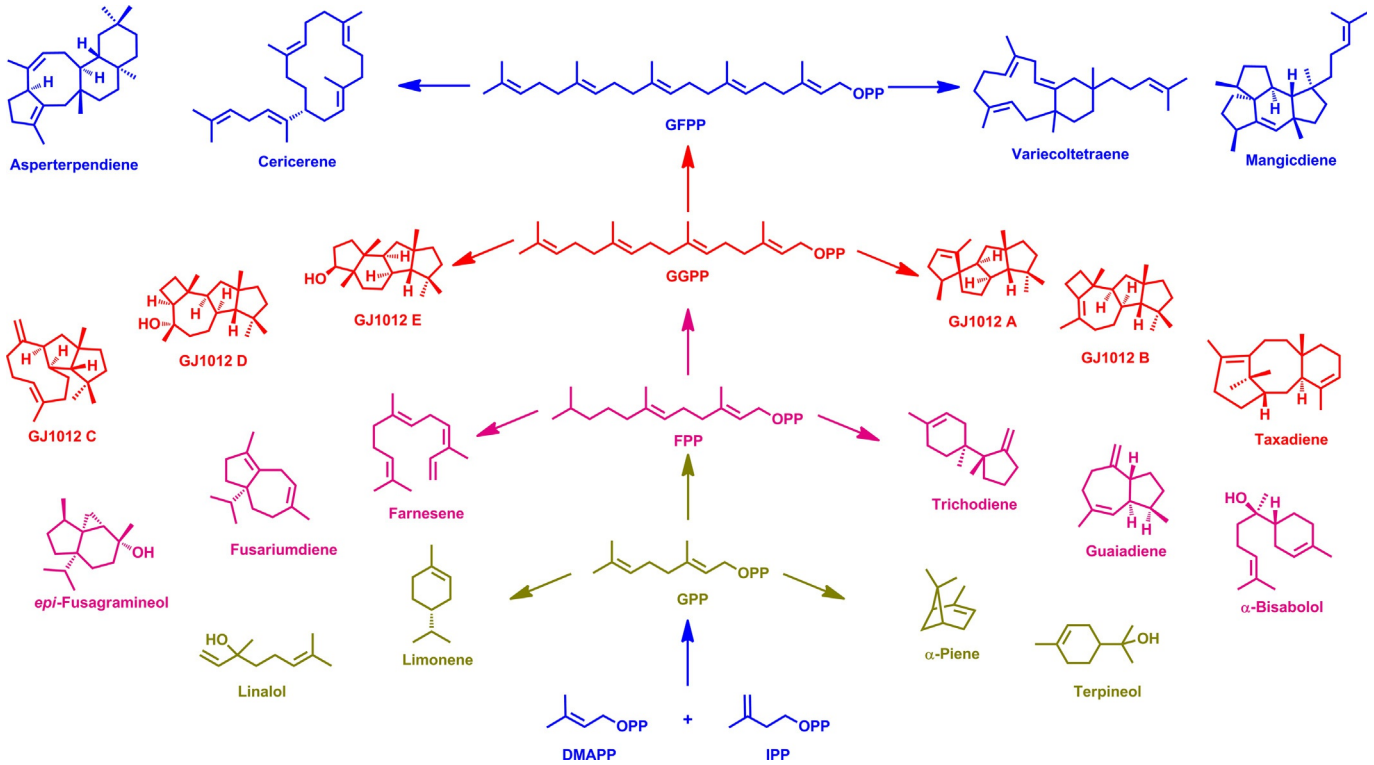

Fig. 7 Diversity of terpenoid biosynthesis. Examples of monoterpenes (C10), sesquiterpenes (C15), diterpenes (C20), and sesterterpenes (C25) are shown.

effects of gene silencing, complex regulatory systems, or low yields in the native host, accelerating the entire process for the mining of terpene cyclases and the discovery of terpenoids (Bian, Han, et al., 2017; Bian et al., 2018; Jia et al., 2016; Shao et al., 2017). Consequently, a metabolic engineering-based terpenoid discovery strategy may accelerate the process of genome mining for terpene cyclases and increase the chemical diversity.

3.1 Genome Mining of Terpene Cyclases and Exploration of the Chemical Diversity of Terpenoids

Two kinds of terpene cyclases exist in nature; class I terpene synthases initiate carbocation formation by substrate ionization and class II terpene cyclases initiate carbocation formation by substrate protonation (Meguro, Tomita, Nishiyama, & Kuzuyama, 2013). Class I terpene cyclases FgMS and FgGS were predicted from *F. graminearum* J1-012 with the highly conserved aspartate-rich motif **DD**VI**E** and the NSE/DTE triad **ND**LY**S**YDK**E**, which chelate Mg^{2+} ions used to bind to the diphosphate moiety of the substrate and required for the ionization of the allylic diphosphate ester bond (Meguro et al., 2013; Shishova, Di Costanzo, Cane, & Christianson, 2007). A bioinformatic analysis showed that FgMS is a chimeric terpene cyclase with the prenyltransferase and the terpene cyclase domain and can utilize IPP and DMAPP as substrates to produce terpene skeletons. In vitro assays have shown that both of these enzymes could directly catalyze the conversion GPP, FPP, GGPP, and GFPP to generate multiple products, exhibiting exceptionally broad substrate specificities for polyisoprenoid diphosphates (Bian, Han, et al., 2017).

Directed by the combinatorial biosynthesis approach, a platform consisting of the terpene-overproducing chassis, the prenyltransferase (FPPS, GGPPS, and GFPPS) library, and the enzymes FgMS and FgGS was built to realize the full potential of these terpene synthases. First, the above-mentioned efficient precursor production platform was employed to provide sufficient IPP and DMAPP for terpenoid overproduction (Zhu et al., 2014). Second, efficient prenyltransferases (FPPS, GGPPS, and GFPPS) were employed to construct a library to produce FPP, GGPP, and GFPP for terpene cyclase to produce sesqui-, di-, and sesterterpenes, respectively. Third, the terpene cyclase (FgMS and FgGS) was combined with appropriate prenyltransferases to generate relevant mutants. In this section, we describe the construction of three mutants for the characterization of FgMS and FgGS and the production of sesqui-, di-, and sesterterpenes.

3.1.1 Identification of Class I Terpene Synthases

Class I terpene synthases were identified by screening all predicted proteins against a set of hidden Markov models constructed for each terpene synthase family. Proteins with the highly conserved aspartate-rich motif **DD**xxD/**E** and the NSE/DTE triad (**N/H)D**xx**S**xxx**E** were selected for subsequent analyses. Based on sequence alignments, both FgMS and FgGS contain the highly conserved aspartate-rich motif **DD**xxD/**E** and NSE/DTE triad (Fig. 8), which make it possible to produce terpene skeletons using an appropriate polyisoprenoid diphosphate as a substrate.

3.1.2 Construction of Plasmids

Primers used for plasmid construction are listed in Table 1. Codon-optimized FgMS was amplified using the primer pair P41/P42 and subcloned into pET28a (+) to generate pGB301. The coding sequence of FgGS was assembled and then amplified using the primer pair P43/P44 and subcloned into pET28a (+) and pET21a (+) to generate pGB303 and pGB304, respectively. *FPPS* was amplified using the primer pair P45/P46 and subcloned into pET21a to generate pGB305. Next, *idi* was amplified using the primers P47/P48 and subcloned into pET21a (+) and pETduet-1 to generate pGB306 and pGB307, respectively. Then, pGB307 was digested with *Xba*I/*Xho*I, and the *idi* fragment was ligated into pGB305, which was digested by *Spe*I/*Xho*I, to generate pGB308. The *Xba*I/*Xho*I digested fragment (*FPPS-Idi*) of pGB308 was then inserted into the *Spe*I/*Xho*I digested plasmid pGB304 to generate pGB314. FgMS, GFPPS, and the backbone of pGB307 were amplified using the primer pairs P49/50, P51/P52, and P53/P54, respectively, and these three fragments were then assembled by the Gibson method to generate pGB309 (Gibson et al., 2009). Similarly, FgGS, GGPPS, and the backbone of pGB307 were amplified using the primer pairs P55/56, P57/P58, and P53/P59 and then assembled to generate pGB313.

3.1.3 In Vitro Assays of Terpene Cyclases

To test the in vitro activities of FgMS and FgGS, the following reactions were conducted.

1. First, 10 μ*M* purified proteins, 100 μ*M* substrates (GPP, FPP, GGPP, or GFPP), and 2 m*M* Mg^{2+} in 200 μL of 50 m*M* PB buffer (pH 7.6) with 10% glycerol were incubated at 30°C overnight.

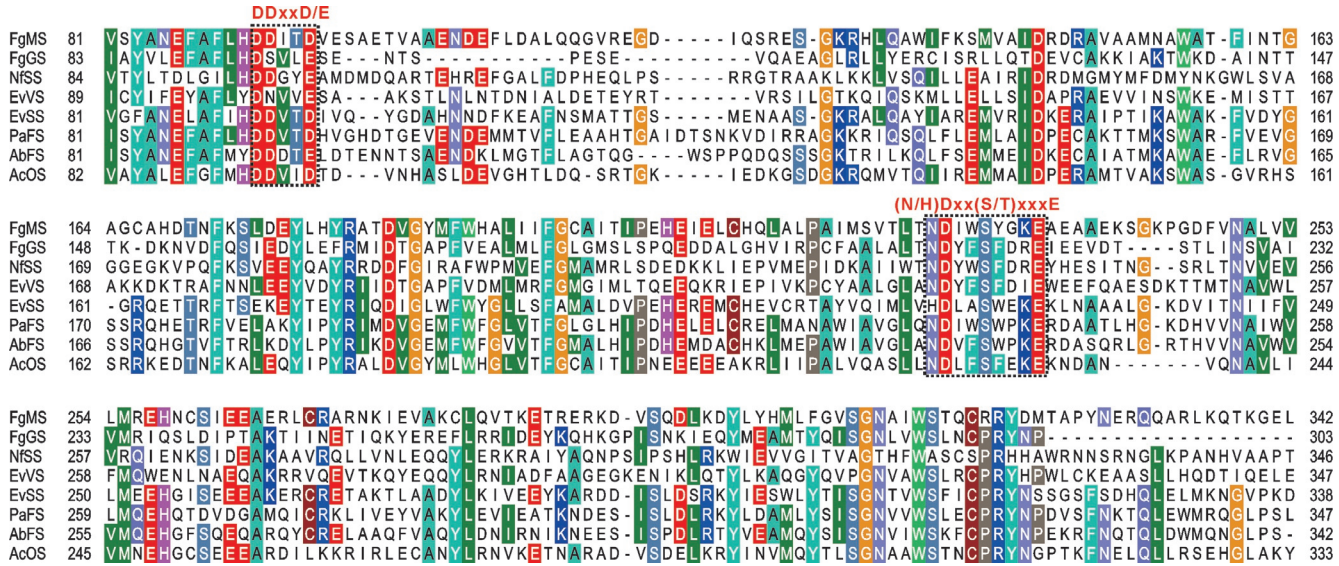

Fig. 8 Sequence alignment of FgMS, FgGS, and the class I terpene synthase of filamentous fungi. The *black boxes* represent the conserved "DDXXD/E" and "NSE/DTE" motifs.

2. The samples were extracted with 200 μL hexane twice and centrifuged at 1000 rpm for 5 min; the organic layer was collected.
3. The production of terpenes was detected GC–MS using the method described in Section 2.1.2.

In the absence of IPP, both FgMS and FgGS could directly catalyze the conversion GPP, FPP, GGPP, and GFPP to generate 51 sesqui-, di-, and sesterterpenes in vitro, as detected by GC–MS, exhibiting exceptionally broad substrate specificities for polyisoprenoid diphosphates (Fig. 9).

3.1.4 Fermentation and Purification of Terpenoids

E. coli T7 was constructed for the production of sesterterpenes by FgMS. *E. coli* T11 and T12 were constructed for the production of diterpenes and sesquiterpenes produced by FgGS (Table 2).

The fermentation and purification of terpenoids produced by FgMS and FgGS were carried out as follows.

1. *E. coli* T7, T11, and T12 were cultivated in 2-L flasks containing 1 L of LB medium according to the description in Section 2.2.2.
2. The cell cultures were extracted twice with an equal volume of hexane and detected by GC/MS according to the description in Section 2.1.2.
3. The organic layer was concentrated with rotary evaporators.
4. The extract was resolved in methanol (with an appropriate ratio DMSO, about 20%), and separation and purification were performed using the XBridge™ Prep C18 column (Waters, 10×250 mm, 5 μm) and prep-HPLC system with acetonitrile and water as the mobile phase.
5. Structures of sesqui-, di-, and sesterterpenes were determined by comparison with authentic standards, comparisons with mass spectra data from the NIST library, and NMR analyses.

The fermentation data showed that, in accordance with the in vitro data, 1 sesquiterpene, 11 diterpenes, and 8 sesterterpenes were produced by *E. coli* T7, T11, and T12, respectively (Table 2). Two sesterterpenes and five novel diterpenes with three unusual skeletons were structurally characterized from three mutants (Fig. 10). Accordingly, the terpenoid overproduction platform and a combinatorial biosynthesis strategy can accelerate the entire process for the discovery and creation of terpenoids and thereby can expand the chemical diversity of terpene skeletons. This strategy can also be extended to other systems to realize the characterization of terpene cyclases with high efficiency (Bian et al., 2018).

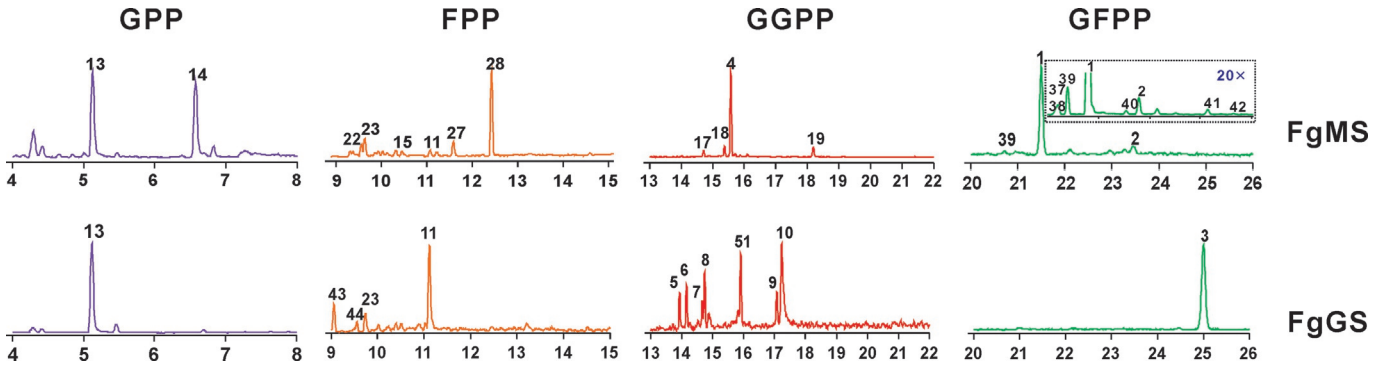

Fig. 9 In vitro characterization of FgMS and FgGS (Bian, Han, et al., 2017). Chromatograms showing the GC–MS analysis of terpenoid products from in vitro assays of FgMS (*top*) and FgGS (*bottom*) with GPP, FPP, GGPP, and GFPP in the absence of IPP.

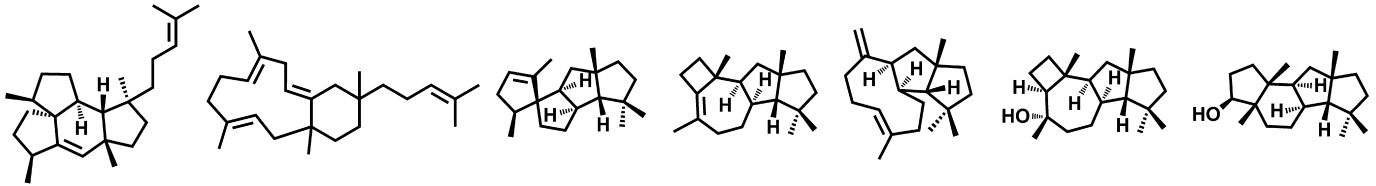

Fig. 10 Structural characterization of seven new terpenoids with three unusual skeletons.

4. CONCLUSION

Terpenoids represent a highly diverse group of natural products with wide applications. The overproduction of value-added terpenoids, the efficient mining of new terpene cyclases, and increased chemical diversity are major goals in the fields of metabolic engineering and natural product discovery. In vitro reconstitution assays provide key information regarding the rate-limiting steps and mechanism of action of the MVA pathway. This information (e.g., the concentration of Idi and the optimal ratio of each component) provides specific direction for engineering research. Directed by the results of these assays, an efficient precursor production platform was constructed, providing a solid foundation for the subsequent overproduction of terpenoids, thereby enabling the overproduction of various kinds of terpenoids, such as farnesene, lycopene, astaxanthin, and taxadiene, with minimal effort and high efficiency.

Metabolic engineering-based genome mining of terpene cyclases makes it possible for researchers to utilize the potential power of terpene cyclases and acquire sufficient quantities of products for structural characterization in a short period of time. In many cases, for promiscuous terpene cyclases that can produce a series of terpenoids, it is only possible to characterize the main products by traditional heterologous or homologous expression, and in vitro feeding experiments are both time-consuming and expensive. The targeted engineering-based precursor production platform can provide a sufficient amount of precursor for the production of mono-, sesqui-, di-, and sesterterpenes with high yields. In this way, the combinatorial biosynthesis approach can release the potential power of terpene cyclases and enable the characterization of all resulting products. As observed in the case of FgGS, some by-products were new terpenes with unusual skeletons, demonstrating that this approach may contribute to increasing the chemical diversity of terpenoids. Owing to the large number of functionally uncharacterized terpene cyclases, an efficient terpenoid precursor-providing platform, the combinatorial biosynthesis approach, the process for genome mining, and the extension of chemical diversity will enable significant advances. This approach is expected to provide a large number of bioactive or high value-added compounds for drug discovery and industrial applications.

ACKNOWLEDGMENTS

This work was financially supported by the funding from J1 Biotech Co., Ltd. Grants from the Young Talents Program of National High-level Personnel of Special Support Program (The "Ten Thousand Talent Program") and the China Postdoctoral Science Foundation (Grant 2017M622507, 2018T110793).

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.

Alper, H., Jin, Y.-S., Moxley, J., & Stephanopoulos, G. (2005). Identifying gene targets for the metabolic engineering of lycopene biosynthesis in *Escherichia coli*. *Metabolic Engineering*, *7*, 155–164.

Andersen-Ranberg, J., Kongstad, K. T., Nielsen, M. T., Jensen, N. B., Pateraki, I., Bach, S. S., et al. (2016). Expanding the landscape of diterpene structural diversity through stereochemically controlled combinatorial biosynthesis. *Angewandte Chemie International Edition*, *55*, 2142–2146.

Bian, G., Han, Y., Hou, A., Yuan, Y., Liu, X., Deng, Z., et al. (2017). Releasing the potential power of terpene synthases by a robust precursor supply platform. *Metabolic Engineering*, *42*, 1–8.

Bian, G., Hou, A., Yuan, Y., Hu, B., Cheng, S., Ye, Z., et al. (2018). Metabolic engineering-based rapid characterization of a sesquiterpene cyclase and the skeletons of Fusariumdiene and Fusagramineol from *Fusarium graminearum*. *Organic Letters*, *20*, 1626–1629.

Bian, G., Yuan, Y., Tao, H., Shi, X., Zhong, X., Han, Y., et al. (2017). Production of taxadiene by engineering of mevalonate pathway in *Escherichia coli* and endophytic fungus *Alternaria alternata* TPF6. *Biotechnology Journal*, *12*, 1600697.

Coussement, P., Bauwens, D., Maertens, J., & De Mey, M. (2016). Direct combinatorial pathway optimization. *ACS Synthetic Biology*, *6*, 224–232.

Fraser, P. D., & Bramley, P. M. (2004). The biosynthesis and nutritional uses of carotenoids. *Progress in Lipid Research*, *43*, 228–265.

Gibson, D. G., Young, L., Chuang, R. Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*, 343–345.

Huang, A. C., Kautsar, S. A., Hong, Y. J., Medema, M. H., Bond, A. D., Tantillo, D. J., et al. (2017). Unearthing a sesterterpene biosynthetic repertoire in the *Brassicaceae* through genome mining reveals convergent evolution. *Proceedings of the National Academy of Sciences of the United States of America, 114*, 201705567. https://doi.org/10.1073/pnas.1705567114.

Hundle, B., Alberti, M., Nievelstein, V., Beyer, P., Kleinig, H., Armstrong, G., et al. (1994). Functional assignment of *Erwinia herbicola* Eho10 carotenoid genes expressed in *Escherichia coli*. *Molecular and General Genetics*, *245*, 406–416.

Jia, M., Potter, K. C., & Peters, R. J. (2016). Extreme promiscuity of a bacterial and a plant diterpene synthase enables combinatorial biosynthesis. *Metabolic Engineering*, *37*, 24–34.

Joint FAO/WHO Expert Committee on Food Additives. (2006). *Compendium of food additive specifications: Joint FAO/WHO Expert Committee on Food Additives, 67th meeting 2006*. Food and Agriculture Organization.

Khaldi, N., Seifuddin, F. T., Turner, G., Haft, D., Nierman, W. C., Wolfe, K. H., et al. (2010). SMURF: Genomic mapping of fungal secondary metabolite clusters. *Fungal Genetics and Biology*, *47*, 736–741.

Klapschinski, T. A., Rabe, P., & Dickschat, J. S. (2016). Pristinol, a sesquiterpene alcohol with an unusual skeleton from *Streptomyces pristinaespiralis*. *Angewandte Chemie International Edition*, *55*, 10141–10144.

Korman, T. P., Opgenorth, P. H., & Bowie, J. U. (2017). A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nature Communications*, *8*, 15526.

Liu, Z., Zhang, Y., Jia, X., Hu, M., Deng, Z., Xu, Y., et al. (2017). In vitro reconstitution and optimization of the entire pathway to convert glucose into fatty acid. *ACS Synthetic Biology*, *6*, 701–709.

Lowry, B., Walsh, C. T., & Khosla, C. (2015). In vitro reconstitution of metabolic pathways: Insights into nature's chemical logic. *Synlett*, *26*, 1008–1025.

Ma, T., Zhou, Y., Li, X., Zhu, F., Cheng, Y., Liu, Y., et al. (2016). Genome mining of astaxanthin biosynthetic genes from *Sphingomonas* sp. ATCC 55669 for heterologous overproduction in *Escherichia coli*. *Biotechnology Journal*, *11*, 228–237.

Meguro, A., Tomita, T., Nishiyama, M., & Kuzuyama, T. (2013). Identification and characterization of bacterial diterpene cyclases that synthesize the cembrane skeleton. *Chembiochem*, *14*, 316–321.

Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K., et al. (1990). Elucidation of the *Erwinia uredovora* carotenoid biosynthetic pathway by functional analysis of gene products expressed in *Escherichia coli*. *Journal of Bacteriology*, *172*, 6704–6712.

Nishida, Y., Adachi, K., Kasai, H., Shizuri, Y., Shindo, K., Sawabe, A., et al. (2005). Elucidation of a carotenoid biosynthesis gene cluster encoding a novel enzyme, 2, 2′-β-hydroxylase, from *Brevundimonas* sp. strain SD212 and combinatorial biosynthesis of new or rare xanthophylls. *Applied and Environmental Microbiology*, *71*, 4286–4296.

Paddon, C. J., Westfall, P. J., Pitera, D. J., Benjamin, K., Fisher, K., McPhee, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, *496*, 528–532.

Pechous, S. W., & Whitaker, B. D. (2004). Cloning and functional expression of an (*E*, *E*)-α-farnesene synthase cDNA from peel tissue of apple fruit. *Planta*, *219*, 84–94.

Pemberton, T. A., Chen, M., Harris, G. G., Chou, W. K., Duan, L., Köksal, M., et al. (2017). Exploring the influence of domain architecture on the catalytic function of diterpene synthases. *Biochemistry*, *56*, 2010–2023.

Polakowski, T., Stahl, U., & Lang, C. (1998). Overexpression of a cytosolic hydroxymethylglutaryl-CoA reductase leads to squalene accumulation in yeast. *Applied Microbiology and Biotechnology*, *49*, 66–71.

Qin, B., Matsuda, Y., Mori, T., Okada, M., Quan, Z., Mitsuhashi, T., et al. (2015). An unusual chimeric diterpene synthase from *Emericella variecolor* and its functional conversion into a sesterterpene synthase by domain swapping. *Angewandte Chemie International Edition*, *54*, 1–5.

Schnurr, G., Schmidt, A., & Sandmann, G. (1991). Mapping of a carotenogenic gene cluster from *Erwinia herbicola* and functional identification of six genes. *FEMS Microbiology Letters*, *78*, 157–161.

Shao, J., Chen, Q.-W., Lv, H.-J., He, J., Liu, Z.-F., Lu, Y.-N., et al. (2017). (+)-Thalianatriene and (−)-retigeranin B catalyzed by sesterterpene synthases from *Arabidopsis thaliana*. *Organic Letters*, *19*, 1816–1819.

Shishova, E. Y., Di Costanzo, L., Cane, D. E., & Christianson, D. W. (2007). X-ray crystal structure of aristolochene synthase from *Aspergillus terreus* and evolution of templates for the cyclization of farnesyl diphosphate. *Biochemistry*, *46*, 1941–1951.

Sun, T., Miao, L., Li, Q., Dai, G., Lu, F., Liu, T., et al. (2014). Production of lycopene by metabolically-engineered *Escherichia coli*. *Biotechnology Letters*, *36*, 1515–1522.

Tominaga, K., Hongo, N., Karato, M., & Yamashita, E. (2012). Cosmetic benefits of astaxanthin on humans subjects. *Acta Biochimica Polonica*, *59*.

Wang, C., Yoon, S.-H., Jang, H.-J., Chung, Y.-R., Kim, J.-Y., Choi, E.-S., et al. (2011). Metabolic engineering of *Escherichia coli* for α-farnesene production. *Metabolic Engineering*, *13*, 648–655.

Wani, M. C., Taylor, H. L., Wall, M. E., Coggon, P., & McPhail, A. T. (1971). Plant antitumor agents. VI. Isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifolia*. *Journal of the American Chemical Society*, *93*, 2325–2327.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., et al. (2015). antiSMASH 3.0—A comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, *43*, 237–243.

Xie, W., Lv, X., Ye, L., Zhou, P., & Yu, H. (2015). Construction of lycopene-overproducing *Saccharomyces cerevisiae* by combining directed evolution and metabolic engineering. *Metabolic Engineering*, *30*, 69–78.

Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D. E., et al. (2015). Terpene synthases are widely distributed in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 857–862.

Yang, Y. l., Zhang, S., Ma, K., Xu, Y., Tao, Q., Chen, Y., et al. (2017). Discovery and characterization of a new family of diterpene cyclases in bacteria and fungi. *Angewandte Chemie International Edition*, *56*, 4749–4752.

Ye, Y., Minami, A., Mandi, A., Liu, C., Taniguchi, T., Kuzuyama, T., et al. (2015). Genome mining for sesterterpenes using bifunctional terpene synthases reveals a unified intermediate of di/sesterterpenes. *Journal of the American Chemical Society*, *137*, 11846–11853.

You, C., Zhang, X. Z., & Zhang, Y. H. P. (2012). Simple cloning via direct transformation of PCR product (DNA Multimer) to *Escherichia coli* and *Bacillus subtilis*. *Applied and Environmental Microbiology*, *78*, 1593–1595.

Yu, X., Liu, T., Zhu, F., & Khosla, C. (2011). In vitro reconstitution and steady-state analysis of the fatty acid synthase from *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 18643–18648.

Zhou, P., Ye, L., Xie, W., Lv, X., & Yu, H. (2015). Highly efficient biosynthesis of astaxanthin in *Saccharomyces cerevisiae* by integration and tuning of algal crtZ and bkt. *Applied Microbiology and Biotechnology*, *99*, 8419–8428.

Zhu, F., Lu, L., Fu, S., Zhong, X., Hu, M., Deng, Z., et al. (2015). Targeted engineering and scale up of lycopene overproduction in *Escherichia coli*. *Process Biochemistry*, *50*, 341–346.

Zhu, F., Zhong, X., Hu, M., Lu, L., Deng, Z., & Liu, T. (2014). In vitro reconstitution of mevalonate pathway and targeted engineering of farnesene overproduction in *Escherichia coli*. *Biotechnology and Bioengineering*, *111*, 1396–1405.