J Ind Microbiol Biotechnol
DOI 10.1007/s10295-017-1967-3

FERMENTATION, CELL CULTURE AND BIOENGINEERING - ORIGINAL PAPER

A homomeric geranyl diphosphate synthase-encoding gene from *Camptotheca acuminata* and its combinatorial optimization for production of geraniol in *Escherichia coli*

Lixia Yang[1,3] · Liangzhen Jiang[1] · Wei Li[1,3] · Yun Yang[1,3] · Guolin Zhang[1] · Yinggang Luo[1,2]

Received: 21 February 2017 / Accepted: 7 July 2017

Abstract Geranyl diphosphate (GPP), the unique precursor for all monoterpenoids, is biosynthesized from isopentenyl diphosphate and dimethylallyl diphosphate via the head-to-tail condensation reaction catalyzed by GPP synthase (GPPS). Herein a homomeric *GPPS* from *Camptotheca acuminata*, a camptothecin-producing plant, was obtained from 5′- and 3′-rapid amplification of cDNA ends and subsequent overlap extension and convenient PCR amplifications. The truncate *CaGPPS* was introduced to replace *ispA* of pBbA5c-MevT(CO)-MBIS(CO, *ispA*), a de novo biosynthetic construct for farnesyl diphosphate generation, and overexpressed in *Escherichia coli*, together with the truncate geraniol synthase-encoding gene from *C. acuminata* (t*CaGES*), to confirm CaGPPS-catalyzed reaction in vivo. A 24.0 ± 1.3 mg L^{-1} of geraniol was produced in the recombinant *E. coli*. The production of GPP was also validated by the direct UPLC-$HRMS^E$ analyses. The t*CaGPPS* and t*CaGES* genes with different copy numbers were introduced into *E. coli* to balance their catalytic potential for high-yield geraniol production. A 1.6-fold increase of geraniol production was obtained when four copies of t*CaGPPS* and one copy of t*CaGES* were introduced into *E. coli*. The following fermentation conditions optimization, including removal of organic layers and addition of new *n*-decane, led to a 74.6 ± 6.5 mg L^{-1} of geraniol production. The present study suggested that the gene copy number optimization, i.e., the ratio of tCaGPPS and tCaGES, plays an important role in geraniol production in the recombinant *E. coli*. The removal and addition of organic solvent are very useful for sustainable high-yield production of geraniol in the recombinant *E. coli* in view of that the solubility of geraniol is limited in the fermentation broth and/or *n*-decane.

Keywords Geranyl diphosphate synthase · *Camptotheca acuminata* · Geraniol · Metabolic engineering · Fermentation

Introduction

Geranyl diphosphate (GPP, Fig. S1, Supplementary Materials), a ten-carbon containing compound, is biosynthesized from the universal five-carbon precursor isopentenyl diphosphate (IPP, Fig. S1, Supplementary Materials) and its isomer dimethylallyl diphosphate (DMAPP, Fig. S1, Supplementary Materials) via the head-to-tail condensation reaction that is catalyzed by homomeric and/or heteromeric GPP synthase (GPPS, Fig. S1, Supplementary Materials) [10, 19]. The above-mentioned five-carbon units are originated from the plastidial 2-*C*-methylerythritol 4-phosphate (MEP) pathway and/or the cytoplasmic mevalonic acid (MVA) pathway [9, 19]. GPP is the committed precursor for all monoterpenes and their derivatives, for instance, geraniol, an acyclic monoterpene alcohol, is a hydrolysis

Electronic supplementary material The online version of this article (doi:10.1007/s10295-017-1967-3) contains supplementary material, which is available to authorized users.

✉ Yinggang Luo
yinggluo@cib.ac.cn

[1] Center for Natural Products Research, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610041, People's Republic of China

[2] State Key Laboratory of Bioorganic and Natural Products Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, People's Republic of China

[3] University of Chinese Academy of Sciences, Beijing 100049, People's Republic of China

Published online: 10 July 2017

product of GPP (Fig. S1, Supplementary Materials). As a commercially important small alcohol, geraniol is widely used in the flavor, fragrance, and energy industries [6, 15, 26, 27, 45].

In our previous report [5], a geraniol synthase-encoding gene from *Camptotheca acuminata*, *CaGES*, was cloned, overexpressed in *Escherichia coli*, and functionally characterized in vitro and in vivo. The truncate CaGES was employed to produce geraniol in *E. coli* via the integration of the encoding genes involved in the whole MVA biosynthetic pathway for generation of IPP and DMAPP and ispA(S80F), a mutant of farnesyl diphosphate (FPP) synthase, that can produce GPP. The geraniol was accumulated to 48.5 ± 0.9 mg L^{-1} under the mentioned experimental conditions [5]. Herein we report the molecular cloning, heterologous overexpression, and functional characterization of a homomeric GPPS-encoding gene from *C. acuminata*, a camptothecin-producing plant in which GPP is the biosynthetic precursor of geraniol, secologanin, and subsequent camptothecin (Fig. S1, Supplementary Materials). The open reading frame of *CaGPPS* was obtained from 5′- and 3′-rapid amplification of cDNA ends and subsequent overlap extension and convenient PCR amplifications. The heterologous overexpression of truncate *CaGPPS*, together with t*CaGES* and the encoding genes of the entire de novo biosynthetic pathway for FPP generation in which *ispA*, an FPP synthase-encoding gene, was replaced by t*CaGPPS*, in *E. coli* led to a 24.0 ± 1.3 mg L^{-1} of geraniol production in the fermentation broth, which confirmed the catalytic potential of tCaGPPS in vivo. The production of GPP was also confirmed by direct UPLC-$HRMS^E$ analyses of the fermentation broth. To optimize the gene copy numbers of t*CaGPPS* and t*CaGES* for high-yield geraniol production, t*CaGPPS* and t*CaGES* with different copies were introduced into *E. coli* and the resultant recombinant strains were fermented. A 1.6-fold of geraniol production was increased when four copies of t*CaGPPS* and one copy of t*CaGES* were introduced into *E. coli*. The following fermentation conditions optimization led to a 3.1-fold increase of geraniol production, i.e., 74.6 ± 6.5 mg L^{-1}, comparing with that of the original recombinant *E. coli*. The present study suggested that the ratio of tCaGPPS and tCaGES plays an important role in high-yield geraniol production in the engineered *E. coli*. Also removal of organic layer and addition of new *n*-decane are very useful for sustainable and high-yield production of geraniol in the engineered *E. coli* in view of that the solubility of geraniol is limited in the fermentation broth and/or *n*-decane.

Materials and methods

Plant materials, seedling growth, and total RNA isolation

The *C. acuminata* seeds collection, seedling growth, and total RNA isolation were performed according to the reported procedures [28].

Rapid amplification of cDNA ends (RACE) of CaGPPS

The gene-specific primers (Table 1) were designed on the basis of the nucleotide sequence of Caa_locus_25448 to amplify the 5′- and 3′-ends of putative *CaGPPS* using 5′- and 3′-RACE methods, respectively. All primers (Table 1) were synthesized, purified, and authenticated by TsingKe Biological Technology (Chengdu) Co., Ltd. The 5′-RACE-Ready-cDNA and 3′-RACE-first strand cDNA mixtures were prepared according to the manual instructions of the manufactures and the reported procedures [5]. For 5′-RACE of *CaGPPS*, a touchup PCR amplification of the 5′-RACE-Ready-cDNA was performed using CaGPPS-5′GSP1 and

Table 1 List of primers used in the study

Primer code	Sequence (5′–3′)	Direction	Application
CaGPPS-5′GSP1	GATGTGCCTATGAAATCAAGAACGTCGTC	Reverse	5′-RACE
CaGPPS-5′GSP2	GTATCTGCATCATCCAGTACATCATCATG	Reverse	
CaGPPS-3′GSP1	TCACAGAGATGATCCATGTGGCTAG	Forward	3′-RACE
CaGPPS-3′GSP2	GACGACGTTCTTGATTTCATAGGCACATC	Forward	
CaGPPS-F	ATGTTATTTTATAGGGGACTTTCTCGG	Forward	Full-length cloning
CaGPPS-R	TTACTTTGTTCTTGTAATTACTCTGTGAGTTAGAC	Reverse	
CaGPPS-BS-F	CGGGATCCGATGGTAGTTGCTGAGGT	Forward	Overexpression
CaGPPS-BS-R	ACGCGTCGACTTACTTTGTTCTTGTAATTAC	Reverse	
pBb-HB-F1	CCCAAGCTTGTGTCCTAGTTGTCAGCG	Forward	Cloning of fragment 1
pBb-HB-R1	ACCATTTCTTTATCCTCCTAGATCCTTATTTAAG	Reverse	
pBb-HB-F2	AAGAAATGGTAGTTGCTGAGGTCCCAAAACTTGCATC	Forward	Site-mutagenesis
pBb-HB-R2	CGCGGATCCTTACTTTGTTCTTGTAATTACTCTG	Reverse	

universal primer A mix as primers and a HiFi Taq DNA polymerase [TransGen Biotech (Beijing) Co., Ltd, China] to afford the 5′-end of *CaGPPS*, following the cycling conditions: 1 cycle of 94 °C for 3 min; 5 cycles of 94 °C for 30 s, 50 °C for 30 s, 72 °C for 1 min; 5 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 1 min; 5 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min; 20 cycles of 94 °C for 30 s, 65 °C for 30 s, 72 °C for 1 min followed by a final extension at 72 °C for 10 min in a thermal cycler (Eppendorf AG, Hamburg, Germany). The PCR products were gel-purified and used as template for the second-round touchup PCR amplification using CaGPPS-5′GSP2 and nested universal primer A as primers, following the identical cycling conditions of the first-round touchup PCR amplification. The amplification products were gel-purified and ligated into the pGM-T vector [Tiangen Biotech (Beijing) Co., Ltd, China], according to the manufacture's instruction. The *E. coli* DH5α competent cells were chemically transformed with the constructs and the nucleotide sequence of the 5′-end of *CaGPPS* was sequenced in TsingKe Biological Technology (Chengdu) Co., Ltd.

The 3′-RACE-first strand cDNA mixtures were used as template, the CaGPPS-3′GSP1 and universal primer A mix as primers, and a HiFi Taq DNA polymerase [TransGen Biotech (Beijing) Co., Ltd, China] to get the 3′-end of CaGPPS, following the touchdown PCR cycling conditions: 1 cycle of 94 °C for 3 min; 5 cycles of 94 °C for 30 s, 65 °C for 30 s, 72 °C for 50 s; 5 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 50 s; 5 cycles of 94 °C for 30 s, 58 °C for 30 s, 72 °C for 50 s; 20 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 50 s followed by a final extension at 72 °C for 10 min. The amplification products were gel-purified and used as template for the second-round PCR amplification with CaGPPS-3′GSP2 and CaGPPS-R as primers, following the touchdown PCR program: 1 cycle of 94 °C for 3 min; 5 cycles of 94 °C for 30 s, 62 °C for 30 s, 72 °C for 45 s; 5 cycles of 94 °C for 30 s, 60 °C for 30 s, 72 °C for 45 s; 5 cycles of 94 °C for 30 s, 58 °C for 30 s, 72 °C for 45 s; 20 cycles of 94 °C for 30 s, 50 °C for 30 s, 72 °C for 45 s followed by a final extension at 72 °C for 10 min. The amplification products were gel-purified and ligated into the pGM-T vector. The *E. coli* DH5α competent cells were chemically transformed with the constructs following the procedure mentioned above and the nucleotide sequence of the 3′-end of *CaGPPS* was sequenced in TsingKe Biological Technology (Chengdu) Co., Ltd.

Full-length cloning of coding sequence of *CaGPPS*

Based on the nucleotide sequences of the 5′- and 3′-ends of *CaGPPS*, two primers CaGPPS-F and -R (Table 1) were designed to clone the full coding sequence of *CaGPPS* by using the overlap extension PCR (OEPCR) amplification methodology. Briefly, the 5′- and 3′-end fragments of *CaGPPS* were used as template and followed the cycling conditions: one cycle of 94 °C for 3 min; ten cycles of 94 °C for 30 s, 57 °C for 30 s, 72 °C for 40 s followed by a final extension at 72 °C for 10 min firstly. The CaGPPS-F and -R primers were added to the above prepared mixture and OEPCR amplification was performed by following the cycling conditions: 1 cycle of 94 °C for 3 min; 20 cycles of 94 °C for 30 s, 57 °C for 30 s, 72 °C for 1 min 10 s followed by a final extension at 72 °C for 10 min to afford the full coding sequence of *CaGPPS*. The full coding sequence of *CaGPPS* was also amplified from the *C. acuminata* cDNA by using the CaGPPS-F and -R as primers (Table 1), a HiFi Taq DNA polymerase [TransGen Biotech (Beijing) Co., Ltd, China], and following the cycling conditions: 1 cycle of 94 °C for 3 min; 37 cycles of 94 °C for 30 s, 53 °C for 30 s, 72 °C for 35 s followed by a final extension at 72 °C for 10 min. The amplification products were gel-purified and ligated into the pGM-T vector. The *E. coli* DH5α competent cells were chemically transformed with the constructs and sequenced following the procedure mentioned above. The full coding DNA sequence of *CaGPPS* was deposited at GenBank under accession number KY484908.

Bioinformatic analyses and 3-dimensional structure prediction of CaGPPS

The bioinformatics analyses and 3-dimensional structure prediction of CaGPPS were performed according to the reported procedures [28].

Generation of overexpression plasmids

The CaGPPS-BS-F and -R primers (Table 1) were designed and synthesized to amplify the truncate *CaGPPS* by using a HiFi Taq DNA polymerase [TransGen Biotech (Beijing) Co., Ltd, China] with the following cycling conditions: 1 cycle of 94 °C for 3 min; 25 cycles of 94 °C for 30 s, 57 °C for 30 s, 72 °C for 40 s, followed by a final extension at 72 °C for 10 min. The PCR products were gel-purified, digested with *Bam*HI and *Sal*I endonucleases, and ligated into the pETDuet-1 harboring t*CaGES* [5] that was digested with the same endonucleases to afford the expression construct pETDuet-1-t*CaGPPS*-t*CaGES* (I, Fig. 1a). The expression constructs III, IV, and V (Fig. 1a) were prepared following the same experimental procedures.

The expression construct II (Fig. 1a) was prepared via the replacement of *ispA* of pBbA5c-MevT(CO)-MBIS(CO, *ispA*) (http://www.addgene.org/35151/) by t*CaGPPS*. The primers pBb-HB-F1 and -R1 (Table 1) were designed and synthesized to amplify the 9148–10,350 bp of pBbA5c-MevT(CO)-MBIS(CO,

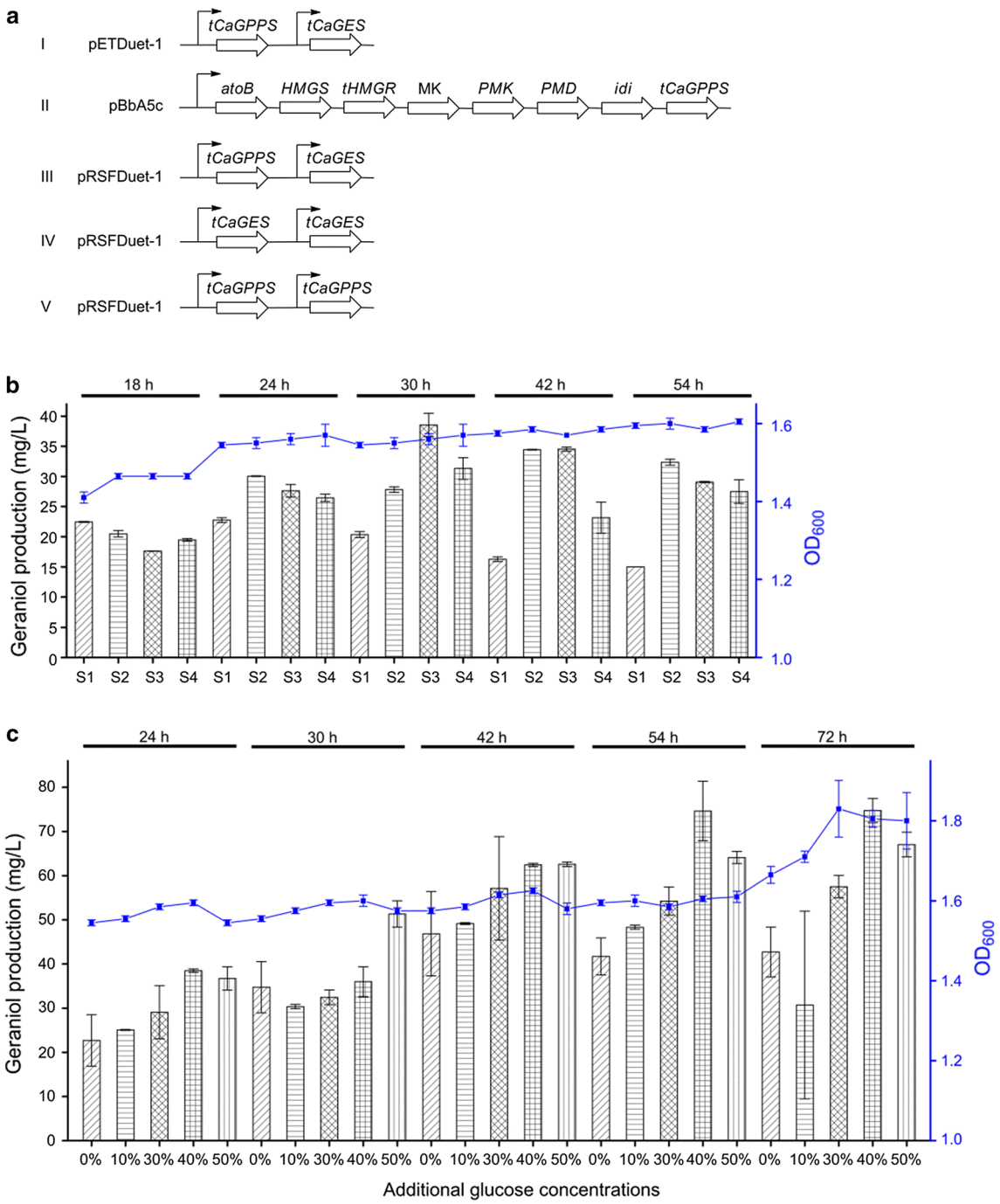

Fig. 1 Geraniol production in *E. coli* using t*CaGPPS* and t*CaGES*. **a** The schematic representation of plasmids I–V. **b** The geraniol yield (*column*) and the biomass (*blue line*) of the corresponding recombinant *E. coli* harboring different plasmids and fermented with different time intervals. The recombinant *E. coli* strains are S1 (2 tCaGPPS + 1 tCaGES), S2 (3 tCaGPPS + 2 tCaGES), S3 (4 tCaGPPS + 1 tCaGES), and S4 (2 tCaGPPS + 3 tCaGES). **c** The effects of fermentation time intervals, the addition of different concentrations of glucose, and replacement of *n*-decane on the geraniol yield (*column*) and the biomass (*blue line*) of recombinant *E. coli*

ispA) and the first five nucleotides ATGGT of 5′-end of t*CaGPPS* to afford Fragment 1 (Fig. S2, Supplementary Materials), following the cycling conditions: 1 cycle of 94 °C for 3 min; 20 cycles of 94 °C for 30 s, 57 °C for 30 s, 72 °C for 40 s followed by a final extension at 72 °C for 10 min. The amplification products were gel-purified and ligated into the pGM-T vector. The *E. coli* DH5α competent cells were chemically transformed with the constructs and sequenced following the procedure mentioned above. An unexpected site for enzymatic digestion

of the endonuclease *Hin*dIII is present in the original DNA sequence of t*CaGPPS*. Thus, the primers pBb-HB-F2 and -R2 (Table 1) were designed and synthesized to get Fragment 2 (Fig. S2, Supplementary Materials). Fragment 2 consisted of AAGAA (10,346–10,350 bp) of pBbA5c-MevT(CO)-MBIS(CO, *ispA*), the coding sequence of t*CaGPPS* with the 24th nucleotide G was mutated into A to avoid the enzymatic digestion of the endonuclease *Hin*dIII by a site-mutagenesis methodology, and GGATCC for *Bam*HI digestion was obtained from the cycling conditions: 1 cycle of 94 °C for 3 min; 20 cycles of 94 °C for 30 s, 57 °C for 30 s, 72 °C for 35 s followed by a final extension at 72 °C for 10 min. The amplification products were gel-purified and ligated into the pGM-T vector. The *E. coli* DH5α competent cells were chemically transformed with the constructs and sequenced following the procedure mentioned above. Within Fragments 1 and 2 in hand, an OEPCR amplification leads to afford Fragment 3 (Fig. S2, Supplementary Materials). Briefly, the Fragments 1 and 2 were used as template and followed the cycling conditions: 1 cycle of 94 °C for 3 min; 10 cycles of 94 °C for 30 s, 67 °C for 30 s, 72 °C for 40 s followed by a final extension at 72 °C for 10 min, firstly. The pBb-HB-F1 and -R2 primers (Table 1) were added to the above prepared mixture and OEPCR amplification was performed by following the cycling conditions: 1 cycle of 94 °C for 3 min; 20 cycles of 94 °C for 30 s, 67 °C for 30 s, 72 °C for 1 min 20 s followed by a final extension at 72 °C for 10 min. The amplification products were gel-purified and ligated into the pGM-T vector. The *E. coli* DH5α competent cells were chemically transformed with the constructs and sequenced following the procedure mentioned above. Then the pGM-T-Fragment 3 was digested with *Hin*dIII/*Bam*HI endonucleases and the resulting Fragment 3 was ligated into pBbA5c-MevT(CO)-MBIS(CO, *ispA*) that was digested with the same endonucleases to afford the expression construct II (Fig. 1a).

Heterologous overexpression of recombinant proteins

The *E. coli* BL21(DE3) competent cells were chemically transformed with the expression plasmid(s) to afford the corresponding recombinant strain. A single colony of the recombinant strain was inoculated into 5 mL of Luria–Bertani broth (LB: tryptone, 10 g L^{-1}; NaCl, 10 g L^{-1}; and yeast extract, 5 g L^{-1}) containing the designated antibiotics from the expression plasmid with appropriate concentrations, for instance, 100 μg mL^{-1} of ampicillin for pETDuet-1 origin and its derivatives and 34 μg mL^{-1} of chloramphenicol for pBbA5c and its relatives. The following incubation and protein overexpression were performed on the basis of the reported experimental procedures [5].

Characterization and quantitation of the product from the CaGPPS-catalyzed reaction

The product characterization and quantitation were performed according to the reported experimental procedures [5] with a minor modification. An Agilent ZORBAXSB-C_{18} analytic column (250 mm × 4.6 mm, 5 μm) was used in the HPLC–DAD analyses of the fermentation products.

The UPLC-$HRMS^E$ analyses were performed on a Waters Acquity I Class UPLC system equipped with an Acquity UPLC HSS T3 column (2.1 mm × 100 mm, 1.8 μm) and a Vion IMS QToF mass spectrometer. The mobile phase consisted of buffer A (25 mM ammonium acetate with 0.1% triethylamine) and buffer B [acetonitrile and 25 mM ammonium acetate (4:1) with 0.1% trimethylamine]. The UPLC separation of the sample was performed at 30 °C with the linear gradient of 5–100% buffer B in 20 min and 100–5% buffer B in 4 min. The flow rate is 0.3 mL min^{-1} and the sample injection volume was 10 μL. The mass spectrometer was detected in negative ion mode.

Fermentation of the engineered *E. coli*

A two-phase extractive fermentation procedure was introduced to ferment the recombinant strains prepared above, according to the reported experimental procedures [5].

Optimization for geraniol production in the engineered *E. coli*

The *E. coli* BL21(DE3) competent cells were chemically transformed with the expression plasmids I + II, I + II + III, I + II + V, and I + II + IV to afford the recombinant strains S1 (2 tCaGPPS + 1 tCaGES), S2 (3 tCaGPPS + 2 tCaGES), S3 (4 tCaGPPS + 1 tCaGES), and S4 (2 tCaGPPS + 3 tCaGES), respectively. The strains were fermented and their geraniol production was quantitated at 18, 24, 30, 42, and 54 h, respectively.

Isopropyl *β-D*-1-thiogalactopyranoside [IPTG, Sangon Biotech (Shanghai) Co., Ltd] was added to the fermentation broth to induce the protein overexpression when OD_{600nm} = 0.52 (O1), 0.93 (O2), 1.24 (O3), and 1.45 (O4), respectively. The strains were fermented with rotation rates at 160 and 200 rpm, respectively, and the geraniol production was quantitated at 18, 24, and 30 h, respectively.

Glucose with the concentrations of 1, 5, 10, and 30% of the glucose concentration of M9 were added to the fermentation broth at 24 and 30 h, respectively. The geraniol production was quantitated at 24, 30, and 42 h, respectively.

Based on the results of the above-mentioned, 10, 30, 40, and 50% glucose were added at 24, 30, 42, and 54 h, respectively. Meanwhile 5 mL of organic layer was removed and then 5 mL of new *n*-decane was added to the fermentation broth at 24, 30, 42, and 54 h, respectively. The geraniol production was quantitated at 24, 30, 42, 54, and 72 h, respectively. All experiments were performed in triplicate and the results were described as mean ± SD.

Results

Molecular cloning of CaGPPS and its bioinformatics properties

Three *CaGPPS* candidates, Caa_locus_12676, Caa_locus_25448, and Caa_locus_55375, were annotated in the transcriptome database of *C. acuminata* (http://medicinalplantgenomics.msu.edu). Multiple amino acid residue sequence alignment of the annotated CaGPPSs retrieved from the database showed that only 23 identical amino acid residues present among these candidate CaGPPSs (highlighted in yellow, Fig. 2). Caa_locus_25448 and Caa_locus_55375 might be two partial fragments of a CaGPPS, considering that both of them contain an overlapped and identical 25-amino acid residue motif. Comparing with the amino acid residue sequences of Caa_locus_25448 and Caa_locus_55375, Caa_locus_12676 contains 50 different amino acid residues (Fig. 2), which indicated that there might be two CaGPPSs present in *C. acuminata*. However, Ca_locus_12676 contains only the first aspartate-rich motif (FARM, DDXXD, highlighted in red, Fig. 2). Caa_locus_25448 containing the FARM and a second highly conserved aspartate-rich motif (SARM, DDXXD, highlighted in green, Fig. 2) might be a unique CaGPPS although a stop code present at position 49 (Fig. 2). The two characteristic DDXXD motifs were found in most, if not all, angiosperm plant GPPSs (Fig. 3) and were suggested to be involved in the binding of the diphosphate substrate and metal ion [38]. Thus, 5′- and 3′-RACE methodologies were employed to clone the 5′- and 3′-ends of putative *CaGPPS* based on the results of multiple sequence alignment (Fig. 2). The coding sequence of *CaGPPS* was obtained from the OEPCR amplification using the 5′- and 3′-end fragments as template. Subsequently the nucleotide sequence of *CaGPPS* (GenBank ID: KY484908) was also cloned from *C. acuminata* cDNA mixture using the CaGPPS-F and -R as primers and confirmed by DNA sequencing.

The 1260-bp open reading frame of *CaGPPS* encodes a 419-amino acid residue and contains the two highly conserved DDXXD motifs in homomeric GPPSs (Fig. 3). While the conserved CXXXC motifs that play an important role in the interaction of the heteromeric GPPSs [29, 35, 37] were not found in CaGPPS (Fig. 3). CaGPPS was predicted to be a stable 46.4 kDa protein with theoretical isoelectric point of 6.2 using the ExPASy online bioinformatics tools (http://www.expasy.org/). Similarity search showed that CaGPPS shares 67.95–79.52% amino acid residue identities with functionally characterized angiosperm plant GPPSs (Fig. 3). Phylogenetic analysis indicated that CaGPPS is a new member of the homomeric GPPS subfamily (Fig. S3, Supplementary Materials). The 3-D structure of CaGPPS (Fig. S4, Supplementary Materials) was

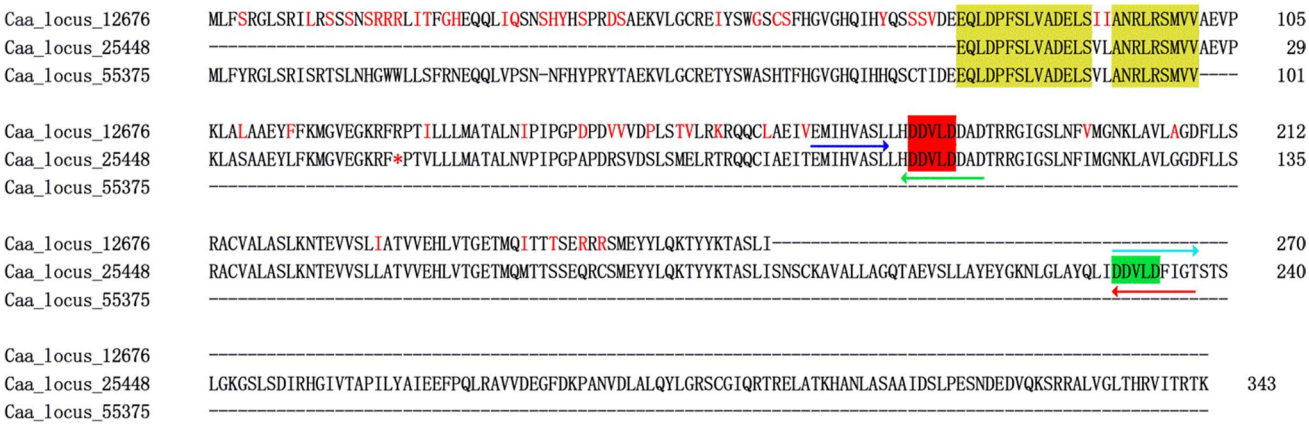

Fig. 2 Multiple amino acid residue sequence alignment of the annotated CaGPPSs from the transcriptome database of *C. acuminata* using Clustal Omega multiple alignment tool. The different amino acid residues among the three annotated CaGPPSs were in *red*, while the identical amino acid residues were in *black*. The amino acid residues highlighted in *yellow* suggested that they are identical within the three annotated CaGPPSs. Those in *red* and *green* indicated the first and the second highly conserved aspartate-rich motifs (DDXXD), respectively. The conserved amino acid residues for 5′-RACE GSP1 and GSP2 primers were indicated in *red* and *green arrows*, respectively. For 3′-RACE, the GSP1 and GSP2 were in *blue* and *cyan arrows*, respectively

Fig. 3 Multiple amino acid residue sequence alignment of functionally characterized GPPSs from angiosperm plants and CaGPPS presented in this study using Clustal Omega alignment tool. The two highly conserved characteristic DDXXD regions were highlighted in *red* and *green*, respectively. The *red arrow* indicated the truncate position for tCaGPPS. The functionally characterized angiosperm plant GPPSs are from *Arabidopsis thaliana* (AtGPPS, CAC16849.1), *Quercus robur* (QrIDS1, CAC20852.1), *Mangifera indica* (MiGPPS, AFJ52721.1), *Azadirachta indica* (AiGDS2, AIG15448.1), *Catharanthus roseus* (CrGPPS, AGL91647.1), and *Solanum lycopersicum* (SlGPPS, ABB88703.1)

predicted and constructed by using the PHYRE2 online server (Protein Homology/analogY Recognition Engine V 2.0, http://www.sbg.bio.ic.ac.uk/phyre2/html/), which suggested that CaGPPS may catalyze the condensation reaction between IPP and DMAPP.

Heterologous overexpression and functional characterization of CaGPPS

Like AtGPPS [2], two methionines Met1 and Met99 (Fig. 3), was predicted to be translation initiation sites in CaGPPS. However, the N-terminal plastidial targeting amino acid residues were truncated to afford the soluble recombinant GPPSs [4]. Thus, the Met1–Ser98 of CaGPPS was truncated to afford soluble tCaGPPS (Fig. 3). In the present work, we propose to characterize the catalytic function of CaGPPS via the replacement of *ispA* (an FPP synthase-encoding gene) of pBbA5c-MevT(CO)-MBIS(CO, *ispA*), a de novo biosynthetic construct for farnesyl diphosphate generation, and co-overexpression with t*CaGES* to validate CaGPPS-catalyzed condensation reaction in vivo. Initially t*CaGPPS* was inserted into the first multiple clone site of the pETDuet-1 harboring t*CaGES* at the second multiple clone site to afford plasmid I (Fig. 1a). To replace the *ispA* gene of pBbA5c-MevT(CO)-MBIS(CO, *ispA*), Fragment 1 was amplified from pBbA5c-MevT(CO)-MBIS(CO, *ispA*) (Fig. S2, Supplementary Materials). Fragment 2 was amplified from t*CaGPPS* with a site-mutagenesis of the 24th nucleotide G into A to avoid the enzymatic digestion of endonuclease *Hind*III (Fig. S2, Supplementary Materials). Subsequent OEPCR amplification and following enzymatic digestion and ligation reactions led to the plasmid II (Fig. 1a). The recombinant strains harboring plasmids I + II and II,

respectively, were grown in LB broth and subsequently fermented in M9 media, comparing with the geraniol-producing *E. coli* strain reported in our previous work [5]. The overexpression of t*CaGPPS* and t*CaGES* was confirmed by SDS-PAGE and western blotting analyses (Fig. S5, Supplementary Materials). HPLC–DAD analyses of the fermentation products showed that only a trace of geraniol was present in the recombinant *E. coli* strain harboring plasmid II only (panel II, Fig. 4), comparing with the standard geraniol (panel IV, Fig. 4) and the product of the *E. coli* strain harboring pBbA5c-MevT(CO)-MBIS[CO, *ispA*(S80F)] and t*CaGES* (panel I, Fig. 4). This might be a very small amount of hydrolysis of GPP to its corresponding alcohol, geraniol, by the *E. coli* strain itself (panel II, Fig. 4). The recombinant *E. coli* strain harboring plasmids I + II affords geraniol (panel III, Figs. 4, S6, Supplementary Materials), which suggested the tCaGPPS-catalyzed condensation reaction between IPP and DMAPP.

According to previous detection methods for isoprenoid diphosphates [21, 23, 30, 40], UPLC-HRMS^E analysis with minor modification was introduced to confirm the GPP production directly. UPLC-HRMS^E analyses showed that a small amount of FPP was observed in the fermentation broth of the *E. coli* harboring empty vector as control (Fig. S7a, Supplementary Materials), which is consistent with previous results [30]. The *E. coli* harboring plasmid II (Fig. 1a) produced GPP, which was confirmed by the UPLC-HRMS^E analysis (Fig. S7b, Supplementary Materials). A slight decrease of FPP was observed in the fermentation broth of the *E. coli* harboring plasmid II (Fig. S7c, Supplementary Materials).

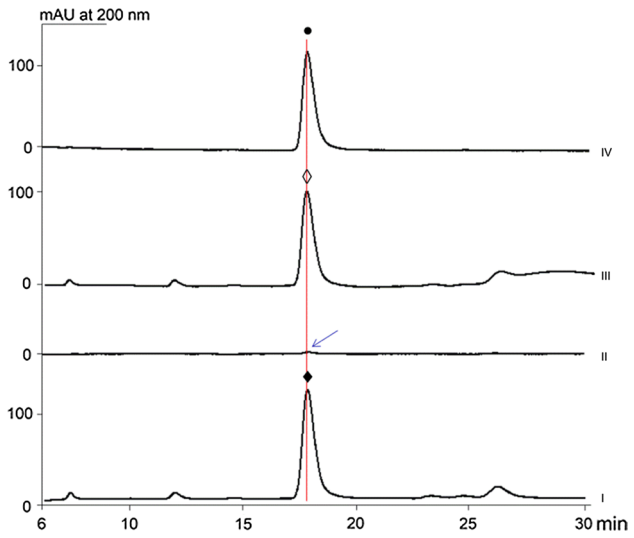

Fig. 4 Functional characterization of the recombinant tCaGPPS. HPLC–DAD analyses of the standard geraniol (*panel IV*), the fermentation product from the *E. coli* harboring pBbA5c-MevT(CO)-MBIS [CO, *ispA*(S80F)] and t*CaGES* (*panel I*) [5], the *E. coli* harboring plasmid II (*panel II*), and the *E. coli* harboring plasmids I + II (*panel III*). *Circle* standard geraniol, *open diamond* product from the *E. coli* harboring plasmids I + II, *filled diamond* product from the *E. coli* harboring pBbA5c-MevT(CO)-MBIS [CO, *ispA*(S80F)] and t*CaGES*

Production of geraniol using tCaGPPS and tCaGES in *E. coli*

Using the optimized fermentation conditions reported previously for tCaGES [5], the recombinant *E. coli* harboring plasmids I + II produced a 24.0 ± 1.3 mg L^{-1} of geraniol that is lower than that reported previously [5]. In view of that both t*CaGPPS* and t*CaGES* were originated from *C. acuminata*, we suppose that the gene copy numbers of t*CaGPPS* and t*CaGES* may have effects on the geraniol production in *E. coli*. Plasmids III, IV, and V (Fig. 1a) were prepared and introduced to afford different recombinant *E. coli* strains. The engineered *E. coli* strains S1, S2, S3, and S4 were fermented and their geraniol production were quantitated at different fermentation times (Fig. 1b). Among the recombinant strains, S3 showed the highest geraniol production at 30 h (Fig. 1b), which indicated the appropriate ratio of the gene copy numbers of t*CaGPPS* and t*CaGES* is 4:1. The IPTG addition time point plays an important role in the geraniol production of the engineered *E. coli* (Fig. S8a and b, Supplementary Materials). The oxygen supply has little effects on the geraniol production when the engineered *E. coli* was fermented for 18–30 h (Fig. S8a and b, supplementary Materials). Thus, the strain S3 was selected as the target strain and IPTG was added when the OD_{600nm} value is −0.7, and fermented with rotation rate at 160 rpm to produce geraniol in *E. coli*.

Considering that the fermentation time for geraniol production in the strain S3 is longer than that reported in the previous work [5], different concentrations of glucose were added to the fermentation broth at different fermentation times to enhance geraniol production (Fig. S8c, Supplementary Materials). A slight increase of geraniol production was observed when 30% glucose was added at 24 h and the fermentation products were harvested at 30 h (Fig. S8c, Supplementary Materials), comparing with that the strain S3 was fermented for 30 h (Fig. 1b). The solubility of geraniol, an acyclic monoterpene alcohol, is limited in the fermentation broth and/or in *n*-decane solvent. Thus, *n*-decane organic layer removal and addition of new *n*-decane were introduced to get rid of the product-feedback inhibition and to alleviate potent toxicity of the product with high concentration (Fig. 1c). The highest geraniol yield was obtained when 40% of glucose was added at 24, 30, 42, and 54 h, the *n*-decane was replaced at 24, 30, 42, and 54 h, and the recombinant strain was fermented for 54–72 h (Fig. 1c). The geraniol production was increased to 74.6 ± 6.5 mg L^{-1} under the optimized fermentation conditions.

Discussion

GPPS catalyzes the head-to-tail condensation reaction between IPP and DMAPP to form a ten-carbon compound, GPP, the universal precursor of all monoterpenoids [10, 19]. As one of the key branch point enzymes of terpene biosynthesis, GPPS is a member of the short-chain *E*-prenyltransferase family [32]. GPPSs have been functionally characterized as homomeric and heteromeric structures in plants [24, 29]. Homomeric GPPSs contain two highly conserved functional aspartate-rich motifs FARM ($DDX_{(2-4)}D$) and SARM (DDXXD) (Fig. S9, Supplementary Materials). Two identical DDVLD motifs were recognized as FARM and SARM in the most of the functionally characterized GPPSs from angiosperm plants (Fig. 3). It should be noted that *Phalaenopsis bellina* lacked the highly conserved FARM and SARM. Instead, a glutamate-rich sequence (EAEVE) was identified in the SARM region (Fig. S9, Supplementary Materials). For heteromeric GPPS, it consists of a large subunit (LSU) that contains the two highly conserved FARM and SARM and a non-catalytic small subunit (SSU) that lacks both motifs [29]. The GPPS.LSU itself might be inactive or function as geranylgeranyl diphosphate synthase. Whereas, an active heteromeric GPPS will be assembled from the interaction between GPPS.LSU and GPPS.SSU [29, 35, 37]. Thus, the characteristic CXXXC motifs are highly conserved in heteromeric GPPSs for the above-mentioned interactions [37]. In the present work, CaGPPS contains the two highly conserved DDVLD motifs for catalytic function and substrate binding and absent of the conserved CXXXC motif characteristic for heteromeric GPPSs (Fig. 3). Phylogenetic analysis showed that CaGPPS belongs to the homomeric GPPS clan (Fig. S3, Supplementary Materials). CaGPPS is a very close relative to CrGPPS (Fig. S3, Supplementary Materials), a GPPS from *Catharanthus roseus*, a well-known monoterpene indole alkaloids producer [29]. Three-dimensional structure prediction suggested that CaGPPS will catalyze the head-to-tail condensation reaction between IPP and DMAPP via a concerted ionization–condensation process [3, 36] (Fig. S4, Supplementary Materials).

Met1 and Met99 of CaGPPS were predicted to be served as potential translation initiation sites (Fig. 3), which suggested that CaGPPS has two forms, similar to AtGPPS [2]. Translation initiation from Met1 will afford a plastid-targeted GPPS isoform, while from Met99, a truncate CaGPPS isoform will be targeted to the cytosol [2]. It should be noted that AtGPPS, a homomeric GPPS from *Arabidopsis thaliana*, was initially characterized to be not involved in high-molecular-weight prenyl diphosphate synthesis [2]. Recently it was suggested to be a polyprenyl pyrophosphate synthase, an unusual prenyltransferase that may produce multiple products with medium/long chain length ranging from 25-carbon to 45-carbon [11]. In the present work, tCaGPPS showed the catalytic activity toward IPP and DMAPP via the geraniol production in the engineered *E. coli*. Other geraniol derivatives and/or polyprenyl pyrophosphate relatives were not present in the fermentation broth (Fig. 4 and Fig. S7, Supplementary Materials).

Metabolic engineering and fermentation optimization have been proven to be valuable methods to enhance yield and accessibility for natural products [5, 14, 22]. Geraniol is a widely used acyclic simple alcohol in many industries [6, 15, 26, 27, 45]. Herein, under the optimized fermentation conditions for geraniol production in *E. coli* using tCaGES [5], a 24.0 ± 1.3 mg L^{-1} of geraniol was produced. The geraniol yield is lower than that reported [5], which suggested that the availability of GPP is limited in the engineered *E. coli*. In view of that the main point for improving a monoterpene product is to improve the availability of GPP by monoterpene synthase [8], gene copy numbers of t*CaGPPS* and t*CaGES* were optimized to balance their catalytic potential for high-yield geraniol production. A 1.6-fold increase of geraniol production was obtained when 4 copies of t*CaGPPS* and 1 copy of t*CaGES* were introduced into *E. coli*. The following fermentation optimizations, including the time for IPTG addition, fermentation time, removal of organic layers and addition of new *n*-decane, led to a 74.6 ± 6.5 mg L^{-1} of geraniol production that is higher than the yields of geraniol and other terpenes production in metabolically engineered microorganisms fermented in shaking flasks [16, 20, 31, 39, 41]. The present study suggested that gene copy number optimization, i.e., the ratio of tCaGPPS and tCaGES, plays an important role in geraniol production in the engineered *E. coli*. The removal and addition of organic solvent are very useful for sustainable and high-yield production of geraniol in the engineered *E. coli* in view of that the solubility of geraniol is limited in the fermentation broth and/or *n*-decane. Recently geraniol was reported to be transformed to geranyl acetate by an *E. coli* acetylesterase in a fed-batch culture system [16]. By hydrolyzing geranyl acetate to geraniol via metabolic engineering, the geraniol production was increased from 68.6 mg L^{-1} in shake flasks to 2.0 g L^{-1} in fed-batch controlled fermentation conditions [16]. However, no such geranyl acetate or other derivatives was found in the fermentation broth of the present engineered *E. coli*. Thus, the present study will be a choice for large-scale geraniol production in the engineered *E. coli*, together with other metabolic engineering methods, fed-batch fermentation, and alleviating product toxicity [1, 12, 16, 17, 20, 39, 41–44].

GPP, the unique precursor for all monoterpenoids, is also the entry point for the generation of the monoterpene moiety of camptothecin (CAM), a plant-derived antitumor monoterpene indole alkaloid (Fig. S1, Supplementary

Materials). In *C. acuminata*, GPP is subsequently converted into geraniol, 10-hydroxygeraniol, loganin, secologanin, and CAM via multiple enzymatic catalysis reaction steps [28, 29]. CAM-type drugs are clinically used for their efficient anti-cancer effects against a broad band of tumor types [7, 34]. The main plant resources for CAM extraction and purification are *C. acuminata* native to China and *Nothapodytes foetida* grown in India [18]. The plant-derived CAM cannot supply the heavy demand on it. The biosynthetic investigations of CAM [5, 28, 33] will be helpful to enhance the production amount of CAM via plant molecular breeding, metabolic engineering, etc [13, 25, 33].

Acknowledgements The authors wish to acknowledge the financial support in part by the KSCX2-EW-Q-6 and ZHTS-003 projects from the Chinese Academy of Sciences, the Applied and Basic Research Program of Sichuan Province (2015JY0058), and the 21172216 project from the National Natural Science Foundation of China. All authors have agreed to submit this manuscript to the Journal of Industrial Microbiology & Biotechnology.

References

1. Alonso-Gutierrez J, Chan R, Batth TS, Adams PD, Keasling JD, Petzold CJ, Lee TS (2013) Metabolic engineering of *Escherichia coli* for limonene and perillyl alcohol production. Metab Eng 19:33–41
2. Bouvier F, Suire C, d'Harlingue A, Backhaus RA, Camara B (2000) Molecular cloning of geranyl diphosphate synthase and compartmentation of monoterpene synthesis in plant cells. Plant J 24:241–252
3. Brandt W, Bräuer L, Günnewich N, Kufka J, Rausch F, Schulze D, Schulze E, Weber R, Zakharova S, Wessjohann L (2009) Molecular and structural basis of metabolic diversity mediated by prenyldiphosphate converting enzyme. Phytochemistry 70:1758–1775
4. Burke C, Croteau R (2002) Interaction with the small subunit of geranyl diphosphate synthase modifies the chain length specificity of geranylgeranyl diphosphate synthase to produce geranyl diphosphate. J Biol Chem 277:3141–3149
5. Chen F, Li W, Jiang L, Pu X, Yang Y, Zhang G, Luo Y (2016) Functional characterization of a geraniol synthase-encoding gene from *Camptotheca acuminata* and its application in production of geraniol in *Escherichia coli*. J Ind Microbiol Biotechnol 43:1281–1292
6. Chen W, Viljoen AM (2010) Geraniol—a review of a commercially important fragrance material. S Afr J Bot 76:643–651
7. Demain AL, Vaishnav P (2011) Natural products for cancer chemotherapy. Microbiol Biotechnol 4:687–699
8. Deng Y, Sun M, Xu S, Zhou J (2016) Enhanced (*S*)-linalool production by fusion expression of farnesyl diphosphate synthase and linalool synthase in *Saccharomyces cerevisiae*. J Appl Microbiol 121:187–195
9. Eisenreich W, Rohdich F, Bacher A (2001) Deoxyxylulose phosphate pathway to terpenoids. Trends Plant Sci 6:78–84
10. Gao Y, Honzatko RB, Peters RJ (2012) Terpenoid synthase structure: a so far incomplete view of complex catalysis. Nat Prod Rep 29:1153–1175
11. Hsieh F, Chang T, Ko T, Wang AHJ (2011) Structure and mechanism of an *Arabidopsis* medium/long-chain-length prenyl pyrophosphate synthase. Plant Physiol 155:1079–1090
12. Jiang G, Yao M, Wang Y, Zhou L, Song T, Liu H, Xiao W, Yuan Y (2017) Manipulation of GES and ERG20 for geraniol overproduction in *Saccharomyces cerevisiae*. Metab Eng 41:57–66
13. Kutchan TM (1995) Alkaloid biosynthesis—the basis for metabolic engineering of medicinal plants. Plant Cell 7:1059–1070
14. Lange BM, Ahkami A (2013) Metabolic engineering of plant monoterpenes, sesquiterpenes and diterpenes – current status and future opportunities. Plant Biotechnol J 11:169–196
15. Lapczynski A, Bhatia SP, Foxenberg RJ, Letizia CS, Api AM (2008) Fragrance material review on geraniol. Food Chem Toxicol 46:S160–S170
16. Liu W, Xu X, Zhang R, Cheng T, Cao Y, Li X, Guo J, Liu H, Xian M (2016) Engineering *Escherichia coli* for high-yield geraniol production with biotransformation of geranyl acetate to geraniol under fed-batch culture. Biotechnol Biofuels 9:58
17. Liu W, Zhang R, Tan N, Xu X, Cao Y, Xian M, Liu H (2015) Utilization of alkaline phosphatase PhoA in the bioproduction of geraniol by metabolically engineered *Escherichia coli*. Bioengineered 6:288–293
18. Lorence A, Nessler CL (2004) Camptothecin, over four decades of surprising findings. Phytochemistry 65:2735–2749
19. Mahmoud SS, Croteau R (2002) Strategies for transgenic manipulation of monoterpene biosynthesis in plants. Trends Plant Sci 7:366–373
20. Mi J, Becher D, Lubuta P, Dany S, Tusch K, Schewe H, Buchhaupt M, Schrader J (2014) *De novo* production of the monoterpenoid geranic acid by metabolically engineered *Pseudomonas putida*. Microb Cell Fact 13:170
21. Miriyala S, Subramanian T, Panchatcharam M, Ren H, McDermott MI, Sunkara M, Drennan T, Smyth SS, Spielmann HP, Morris AJ (2010) Functional characterization of the atypical integral membrane lipid phosphatase PDP1/PPAPDC2 identifies a pathway for interconversion of isoprenols and isoprenoid phosphates in mammalian cells. J Biol Chem 285:13918–13929
22. Morrone D, Lowry L, Determan MK, Hershey DM, Xu M, Peters RJ (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
23. Nagel R, Gershenzon J, Schmidt A (2012) Nonradioactive assay for detecting isoprenyl diphosphate synthase activity in crude plant extracts using liquid chromatography coupled with tandem mass spectrometry. Anal Biochem 422:33–38
24. Pandreka A, Dandekar DS, Haldar S, Uttara V, Vijayshree SG, Mulani FA, Aarthy T, Thulasiram HV (2015) Triterpenoid profiling and functional characterization of the initial genes involved in isoprenoid biosynthesis in neem (*Azadirachta indica*). BMC Plant Biol 15:214
25. Peng ZH, Hu JJ, Li HS (2010) Application of biotechnology in research of camptothecin. Prog Modern Biomed 10:3175–3177
26. Peralta-Yahya PP, Keasling JD (2010) Advanced biofuel production in microbes. Biotechnol J 5:147–162
27. Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD, Lee TS (2011) Identification and microbial production of a terpene-based advanced biofuel. Nat Commun 2:483. doi:10.1038/ncomms1494
28. Qu X, Pu X, Chen F, Yang Y, Yang L, Zhang G, Luo Y (2015) Molecular cloning, heterologous expression, and functional characterization of an NADPH-cytochrome P450 reductase gene from *Camptotheca acuminata*, a camptothecin-producing plant. PLoS One 10:e0135397
29. Rai A, Smita SS, Singh AK, Shanker K, Nagegowda DA (2013) Heteromeric and homomeric geranyl diphosphate synthases from

Catharanthus roseus and their role in monoterpene indole alkaloid biosynthesis. Mol Plant 6:1531–1549
30. Reiling KK, Yoshikuni Y, Martin VJJ, Newman J, Bohlmann J, Keasling JD (2014) Mono and diterpene production in *Escherichia coli*. Biotechnol Bioeng 87:200–212
31. Sarria S, Wong B, Martín HG, Keasling JD, Peralta-Yahya P (2014) Microbial synthesis of pinene. ACS Synth Biol 3:466–475
32. Schmidt A, Gershenzon J (2008) Cloning and characterization of two different types of geranyl diphosphate synthases from Norway spruce (*Picea abies*). Phytochemistry 69:49–57
33. Shen SH, Liu JY, Hu JQ, Li XH, Wang LL (2011) Advances in studies on biosynthetic pathways of camptothecin and their synthases. Chin Trad Herbal Drugs 42:1862–1868
34. Sirikantaramas S, Asano T, Sudo H, Yamazaki M, Saito K (2007) Camptothecin: therapeutic potential and biotechnology. Curr Pharm Biotechnol 8:196–202
35. Tholl D, Kish CM, Orlova I, Sherman D, Gershenzon J, Pichersky E, Dudareva N (2004) Formation of monoterpenes in *Antirrhinum majus* and *Clarkia breweri* flowers involves heterodimeric geranyl diphosphate synthases. Plant Cell 16:977–992
36. Wallrapp FH, Pan JJ, Ramamoorthy G, Almonacid DE, Hillerich BS, Seidel R, Patskovsky Y, Babbitt PC, Almo SC, Jacobson MP (2013) Prediction of function of the polyprenyl transferase subgroup in the isoprenoid synthase superfamily. Proc Natl Acad Sci USA 110:E1196–E1202
37. Wang G, Dixon RA (2009) Heterodimeric geranyl(geranyl) diphosphate synthase from hop (*Humulus lupulus*) and the evolution of monoterpene biosynthesis. Proc Natl Acad Sci USA 106:9914–9919
38. Wang KC, Ohnuma S (2000) Isoprenyl diphosphate synthases. Biochim Biophys Acta 1529:33–48
39. Yang J, Nie Q, Ren M, Feng H, Jiang X, Zheng Y, Liu M, Zhang H, Xian M (2013) Metabolic engineering of *Escherichia coli* for the biosynthesis of alpha-pinene. Biotechnol Biofuels 6:60
40. Zhang D, Poulter DC (1993) Analysis and purification of phosphorylated isoprenoids by reversed-phase HPLC. Anal Biochem 213:356–361
41. Zhang H, Liu Q, Cao Y, Feng X, Zheng Y, Zou H, Liu H, Yang J, Xian M (2014) Microbial production of sabinene—a new terpene-based precursor of advanced biofuel. Microb Cell Fact 13:20
42. Zhao J, Bao X, Li C, Shen Y, Hou J (2016) Improving monoterpene geraniol production through geranyl diphosphate synthesis regulation in *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 100:4561–4571
43. Zhao J, Li C, Zhang Y, Shen Y, Hou J, Bao X (2017) Dynamic control of ERG20 expression combined with minimized endogenous downstream metabolism contributes to the improvement of geraniol production in *Saccharomyces cerevisiae*. Microb Cell Fact 16:17
44. Zhou J, Wang C, Wang L, Choi E, Kim S (2015) Geranyl diphosphate synthase: an important regulation point in balancing a recombinant monoterpene pathway in *Escherichia coli*. Enzyme Microb Technol 68:50–55
45. Zhou J, Wang C, Yoon S, Choi E, Kim S (2014) Engineering *Escherichia coli* for selective geraniol production with minimized endogenous dehyrogenation. J Biotechnol 169:42–50