Phytochemistry 78 (2012) 20–28

Contents lists available at SciVerse ScienceDirect

Phytochemistry

journal homepage: www.elsevier.com/locate/phytochem

Functional characterization of four sesquiterpene synthases from *Ricinus communis* (Castor bean)

Xinkai Xie [a,1], James Kirby [a], Jay D. Keasling [a,b,c,d,e,*]

[a] California Institute of Quantitative Biosciences (QB3), University of California, Berkeley, CA 94720, United States
[b] Department of Chemical and Biomolecular Engineering, University of California, Berkeley, CA 94720, United States
[c] Bioengineering, University of California, Berkeley, CA 94720, United States
[d] Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States
[e] Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, CA 94608, United States

ARTICLE INFO

Article history:
Received 6 September 2011
Received in revised form 16 February 2012
Available online 27 March 2012

Keywords:
Castor bean
Ricinus communis
Euphorbiaceae
Genome mining
Secondary metabolism
Terpenoids
α-Farnesene
δ-Cadinene
α-Copaene

ABSTRACT

Genome sequence analysis of *Ricinus communis* has indicated the presence of at least 22 putative terpene synthase (TPS) genes, 13 of which appear to encode sesquiterpene synthases (SeTPSs); however, no SeTPS genes have been isolated from this plant to date. cDNAs were recovered for six SeTPS candidates, and these were subjected to characterization in vivo and in vitro. The RcSeTPS candidates were expressed in either *Escherichia coli* or *Saccharomyces cerevisiae* strains with engineered sesquiterpene biosynthetic pathways, but only two (RcSeTPS1 and RcSeTPS7) produced detectable levels of product. In order to check whether the engineered microbial hosts were adequately engineered for sesquiterpene production, a selection of SeTPS genes was chosen from other plant species and demonstrated consistently high sesquiterpene titers. Activity could be demonstrated in vitro for two of the RcSeTPS candidates (RcSeTPS5 and RcSeTPS10) that were not observed to be functional in our microbial hosts. RcSeTPS1 produced two products, (−)-α-copaene and (+)-δ-cadinene, while RcSeTPS7 produced a single product, (E, E)-α-farnesene. Both RcSeTPS5 and RcSeTPS10 produced multiple sesquiterpenes.

1. Introduction

Ricinus communis (Euphorbiaceae) is an economically significant plant, due in large part to its production of castor oil, which is used as a food additive, flavoring, lubricant, cosmetic ingredient, and in traditional medicine to heal wounds. Another important aspect of this plant is that the seed contains an extremely toxic protein, ricin, which is one of the most toxic substances in nature (Knight, 1979). Safety concerns regarding exposure to ricin has led to a prohibition of widespread planting of *R. communis* in the US. The genome sequence of *R. communis* was completed recently, in part to uncover the biosynthetic genes for ricinoleic acid, the major component of castor oil, and ricin (Chan et al., 2010). During this process, a group of gene products were annotated as terpene synthases (TPSs) through multiple gene-prediction programs and homology searches against the sequence database (Chan et al., 2010). TPSs are a group of enzymes that catalyze the formation of low molecular weight terpene metabolites including hemiterpenes (C5), monoterpenes (C10), sesquiterpenes (C15), and diterpenes (C20), which can be further oxidized and modified to generate terpenoids. Currently, more than 60,000 members of the terpenoid family are known, and a wide assortment of structures has been elucidated (http://dnp.chemnetbase.com/).

Interestingly, very few terpenes or terpenoids have been identified from the extracts of *R. communis* (Darmanin et al., 2009; Sitton and West, 1975; Tan et al., 2009). Terpenes, mainly monoterpenes and sesquiterpenes, and terpene derivatives are primary components of the essential oils of many plants. A standard extraction from the aerial parts of *R. communis* yielded only a small amount of essential oil containing three monoterpenes and one sesquiterpene, β-caryophyllene (Darmanin et al., 2009). To date, a total of three diterpene synthases (DiTPSs), including two casbene synthases and one neocembrene synthase, have been characterized (Kirby et al., 2010; Mau and West, 1994), while no SeTPSs have been isolated and characterized from this plant. In general, SeTPSs in plants are localized in the cytosol and catalyze the formation of acyclic, monocyclic, bicyclic, and tricyclic products from the substrate (E, E)-farnesyl diphosphate (E, E)-FPP (1) (Fig. 1) while monoterpene synthases (MoTPSs) and DiTPSs are localized in the plastid (Chen et al., 2011; Kirby and Keasling, 2009). Consequently, MoTPSs and DiTPSs containing N-terminal transit peptides of around 62–64 amino acids that target these proteins to the plastid are longer than

* Corresponding author at: Joint BioEnergy Institute, University of California, 5885 Hollis Street, Fourth Floor, Emeryville, CA 94608, United States. Tel.: +1 510 642 4862; fax: +1 510 495 2630.
E-mail address: keasling@berkeley.edu (J.D. Keasling).
[1] Present address: Codexis, Inc., 200 Penobscot Drive, Redwood City, CA 94063, United States.

0031-9422/$ - see front matter © 2012 Published by Elsevier Ltd.
doi:10.1016/j.phytochem.2012.02.022

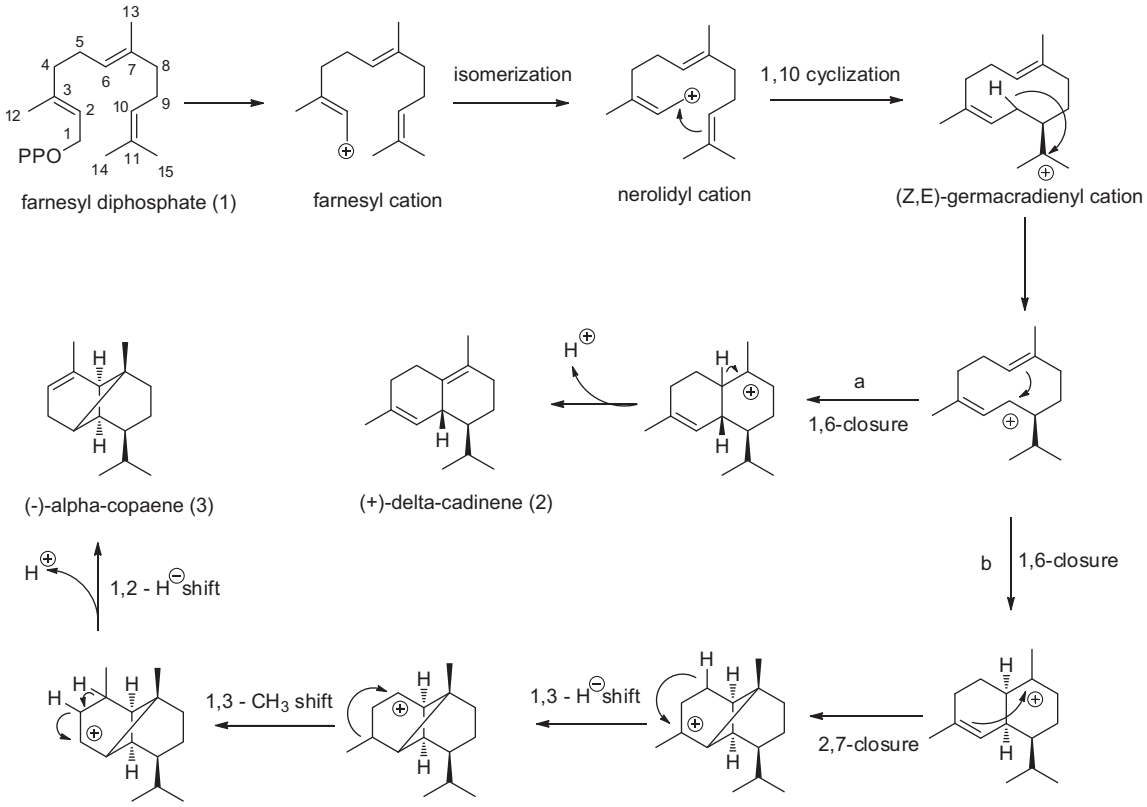

Fig. 1. Proposed cyclization pathway of RcSeTPS1 to form (+)-δ-cadinene (2) (route A) and (−)-α-copaene (3) (route B) from farnesyl diphosphate (1).

SeTPSs containing no signal peptide (Bohlmann et al., 1998; Martin et al., 2004; Williams et al., 1998). In one exception to this rule, two SeTPSs from tomato were localized in the plastid and used (Z, Z)-FPP as substrate to produce sesquiterpenes (Sallaud et al., 2009).

The main purpose of this research was to identify and characterize potential cytosolic SeTPSs in *R. communis* using a genome mining strategy. Four SeTPSs from *R. communis* were functionally characterized, and they were found to produce either single, double, or multiple sesquiterpene products. In vivo production of sesquiterpenes in the engineered microbial hosts using the SeTPSs from *R. communis* was relatively less effective when compared to a random selection of SeTPSs from other plant sources.

2. Results and discussion

2.1. Sequence analysis of SeTPS genes in R. communis and gene isolation

Given that TPSs appear to have evolved independently in different plant lineages (Bohlmann et al., 1998), the sequences of the three characterized DiTPSs from *R. communis* (Kirby et al., 2010; Mau and West, 1994) were used to identify other TPSs in the recently released genome sequence. The predicted SeTPSs from that search were then used to search for other possible SeTPSs. This list was narrowed to a total of 22 TPSs with predicted amino acid lengths of >450, including the three DiTPSs. Since SeTPSs were targeted that are localized to the cytosol and the peptide length of the three plastidic DiTPSs is around 600 amino acids, TPSs with chain length of 540–550 amino acids were predicted to be potential SeTPSs. Allowing for errors in genome sequencing and automated intron prediction, a wider range of candidates were included that ranged in predicted peptide lengths from 476 to 587 amino acids. Thus, from the 22 candidate TPSs, a total of 13 putative SeTPSs were selected for further analysis (Table 1).

In order to obtain the putative SeTPS cDNAs, mRNA was first isolated from fresh leaves of *R. communis* plants, some of which were treated in advance with the elicitor methyl jasmonate (MeJA), and used as template for RT-PCR. Full length cDNAs were amplified using primers designed from the genome sequence of each predicted SeTPS (Table S1). Of the 13 SeTPS candidates, cDNA products for six of them were isolated, while the other seven could not be amplified (Section 4). The seven genes that could not be recovered by RT-PCR may be silent under our growth conditions or could be expressed in tissues that were not tested (Chen et al., 2003; Tholl et al., 2005). Sequencing results from the six putative SeTPS cDNAs established that the predicted peptides generated by genome sequencing and automated annotation were not always accurate (Table 1). The predicted peptide lengths of these six cDNAs range from 547 to 572 amino acids, close to the expected peptide length of a cytosolic SeTPS. Single differences in our predicted peptide sequences were found in two of the SeTPSs, RcTPS1 (T481S, where T is our sequence) and RcSeTPS13 (R205L), relative to the published genome sequence. The peptide lengths of the RcSeTPS5 and RcSeTPS7 cDNAs are 76 and 43 amino acids longer, respectively, than the automated genome sequence predictions.

The encoded proteins of RcSeTPS1, RcSeTPS3, RcSeTPS5, RcSeTPS7, RcSeTPS10, and RcSeTPS13 were aligned (Fig. 2). The aspartate-rich metal binding motif 1, DDXXD, that catalyzes the cleavage of the diphosphate group of the prenyl diphosphate are conserved in all six putative SeTPSs, along with the second metal binding motif, (N,D)D(L,I)X(S,T)XXXE, or a modified version of this, (N,D)D(L,I)X(I)XXXE.

2.2. Characterization of RcSeTPSs in vivo and comparison of product titers to other SeTPSs

To screen the products from the putative SeTPS genes, the six full-length cDNAs were first cloned into expression vectors and

Table 1
Putative SeTPS candidates from *R. communis*.

Name	Protein ID	GenBank accession #	Predicted length[a]	Actual length[b]	Closest characterized homologues (% identity)	Products
RcSeTPS1	XP_002522812	JN315864	558	558	(−)-germacrene D synthase [*Vitis vinifera*] (57%)	(−)-α-copaene, (+)-δ-cadinene
RcSeTPS2	XP_002523635		550			
RcSeTPS3	XP_002523475		547	547		
RcSeTPS4	XP_002513346		563			
RcSeTPS5	XP_002523468	JN315865	476	552	Casbene synthase [*Triadica sebifera*] (43%)	Multiple products
RcSeTPS6	XP_002513369		555			
RcSeTPS7	XP_002518579	JN315866	521	564	Farnesene synthase [*Pyrus communis*] (56%)	(E, E)-α-farnesene
RcSeTPS8	XP_002519897		551			
RcSeTPS9	XP_002515692		584			
RcSeTPS10	XP_002533355	JN315867	551	551	(−)-α-terpineol synthase [*Vitis vinifera*] (69%)	Multiple products
RcSeTPS11	XP_002513266		573			
RcSeTPS12	XP_002534394		587			
RcSeTPS13	XP_002518583		572	572		
	XP_002513334			614	Neocembrene synthase[c]	
	XP_002513340			601	Casbene synthase (CAS1)[c]	
	XP_002513343			599	Casbene synthase (CAS3)[c]	

[a] In amino acids, as annotated from genome sequence.
[b] The predicted peptide length from derived cDNA.
[c] The functions of these DiTPSs were confirmed previously.

expressed in *Escherichia coli* and *Saccharomyces cerevisiae* strains engineered for enhanced isoprenoid production (Redding-Johanson et al., 2011; Ro et al., 2006). These two optimized hosts are capable of producing several hundred milligrams of amorphadiene per liter, primarily as a result of enhanced mevalonate pathway flux to synthesize the precursor (E, E)-FPP (1) (Fig. 1). The purpose of testing production in two different hosts was to circumvent problems associated with codon usage, protein expression, solubility, and activity that may exist in one host but not the other. In this case, however, each of the RcSeTPSs behaved similarly in both hosts, either producing a detectable product or not. Product yields observed from expression in *S. cerevisiae* were higher than the corresponding yields in *E. coli*. RcSeTPS1 produced two sesquiterpene products, and RcSeTPS7 produced a single product (Figs. 3

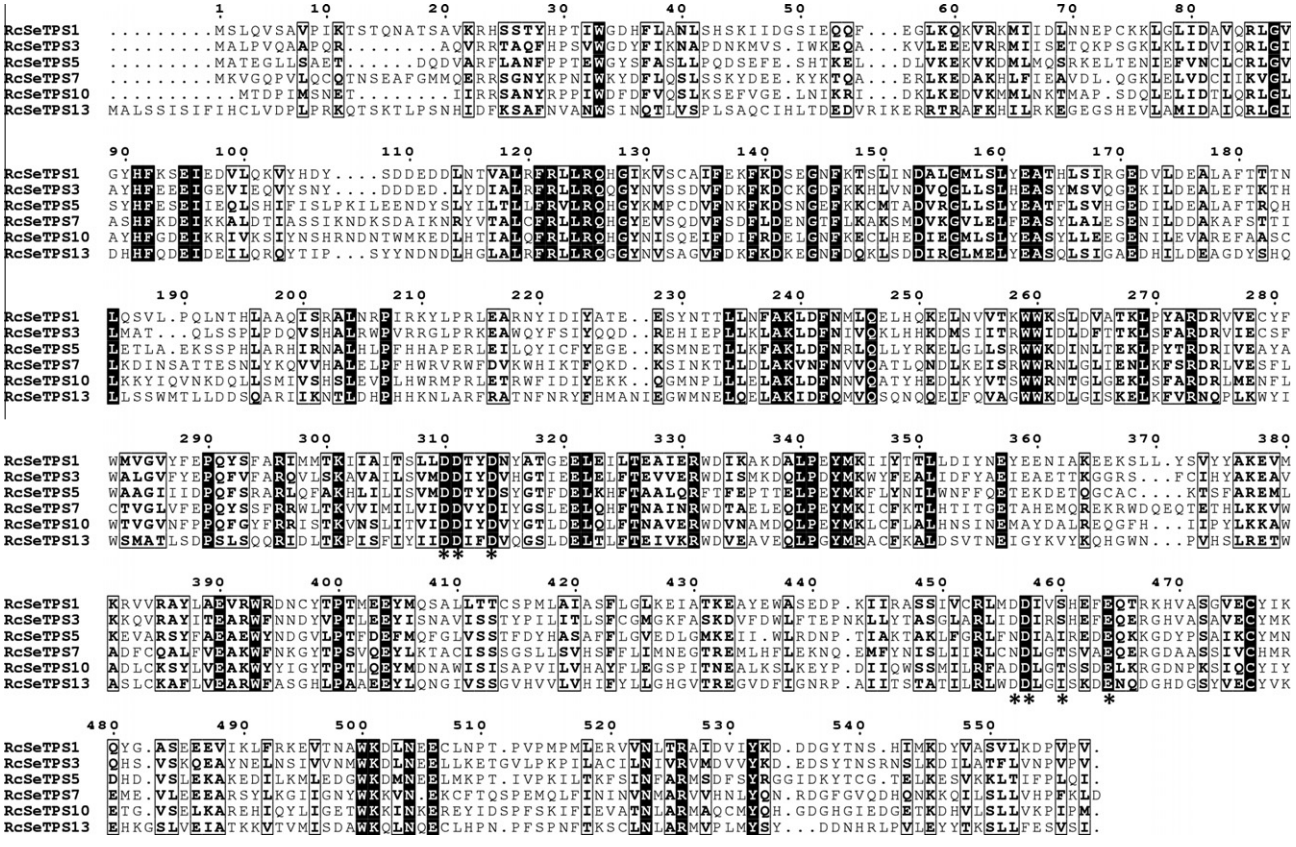

Fig. 2. Sequence alignment of RcSeTPS1, 3, 5, 7, 10, and 13 using ClustalW2. Amino acids completely conserved are highlighted in black boxes. Those with similar properties are enclosed in open boxes. The conserved metal-binding motifs DDXXD and (N,D)D(L,I)X(S, T, I)XXXE are marked with stars.

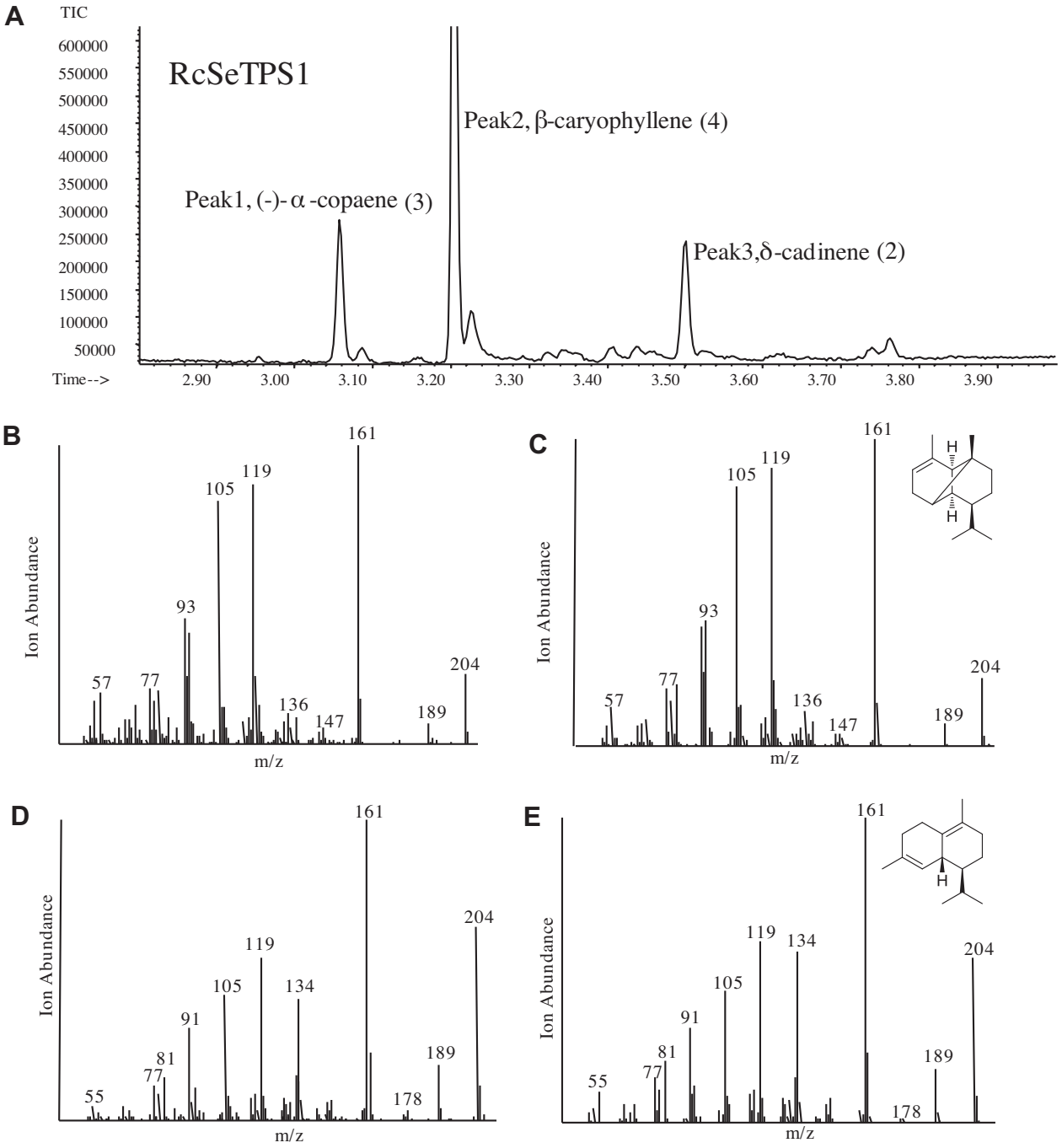

Fig. 3. GC–MS analysis of the sesquiterpene products formed from RcSeTPS1. (A) Total ion count (TIC) chromatogram of two products of RcSeTPS1 together with an internal standard β-caryophyllene (4) for quantification. (B) Mass spectrum of peak 1. (C) Mass spectrum of authentic (−)-α-copaene (3) with the same retention time as peak 1. (D) Mass spectrum of peak 3. (E) Mass spectrum of the product from characterized δ-cadinene synthase, which has the same retention time as peak 3.

and 4). However, expression of the other candidates, namely RcSeTPS3, 5, 10, and 13, failed to yield detectable products in either host (data not shown). The two products of RcSeTPS1 were identified as (−)-α-copaene (3) (~60%), which was confirmed by GC/MS analysis alongside a commercial authentic compound, and (+)-δ-cadinene (2) (~40%), which was confirmed by expressing a characterized cotton (+)-δ-cadinene synthase in the same host and comparing the products by GC-MS (Fig. 3) (Yoshikuni et al., 2006). The titer of (−)-α-copaene (3) at 48 h was 9 mg/L in *S. cerevisiae* and around 7 mg/L in *E. coli*, and the titer of (+)-δ-cadinene (2) was 6 mg/L in *S. cerevisiae* and 3.5 mg/L in *E. coli* (Fig. 3, only shows production in *S. cerevisiae*). The product of RcSeTPS7 was identified as (E, E)-α-farnesene (6), which was confirmed by matching to a major peak in a mixture of farnesenes commercially available, and by searching against the NIST MS database and the Essential Oils HP–Chemstation library (Fig. 4) (Adams, 1995). The product of the characterized (E)-β-farnesene synthase (β-FS) from *Artemisia annua* matches another major peak of the farnesene mixture, which further supports the assignment of (E, E)-α-farnesene (6) to the RcSeTPS7 product (Fig. 4) (Picaud et al., 2005). The titer of (E, E)-α-farnesene (6) was 70 mg/L in *S. cerevisiae* and 12 mg/L in *E. coli* at 48 h. To test if codon usage was the primary reason for the

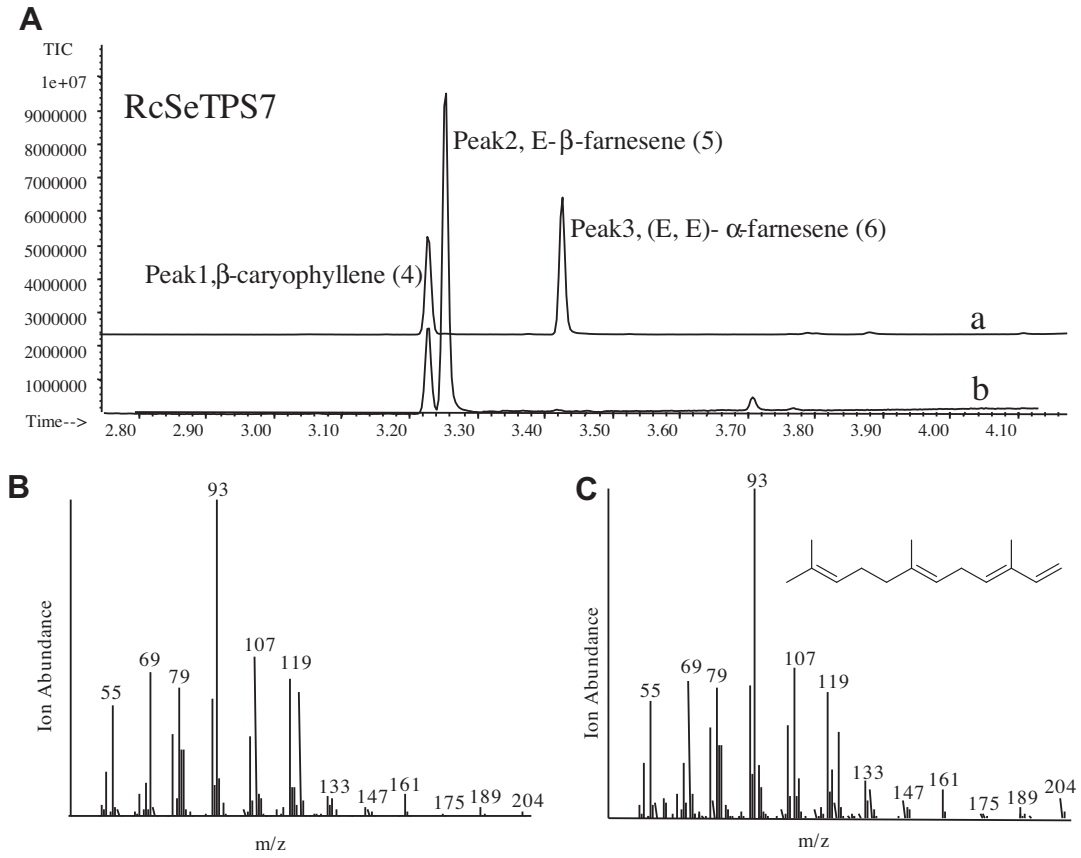

Fig. 4. GC–MS analysis of the sesquiterpene product of RcSeTPS7. (A) Total ion chromatograms of the RcSeTPS7 products mixed with β-caryophyllene (4) as an internal standard for quantification (trace a) and the products from E-β-farnesene synthase also including β-caryophyllene (4) as an internal standard for quantification (trace b). (B) Mass spectrum of peak 3. (C) Mass spectrum of component from the standard farnesene mixture with the same retention time as peak 2.

lack of products observed when the other RcSeTPSs were expressed in our engineered microbial hosts, RcSeTPS3 was subjected to codon optimization and subsequently expressed in *S. cerevisiae* for comparison to the non-codon-optimized cDNA. However, the codon optimized RcSeTPS3 did not produce detectable levels of terpene product either (data not shown).

In order to determine the extent to which our microbial hosts may be limiting the functional expression of SeTPSs and generation of detectable sesquiterpene product levels, several known SeTPSs were randomly selected from other plant species based on availability and ease of access and expressed in our hosts under the same conditions used for the RcSeTPSs. Amorphadiene synthase (ADS) and β-FS from *A. annua* (Martin et al., 2003; Picaud et al., 2005), three germacrene A synthases including AaGAS from *A. annua* (Bertea et al., 2006) and LTC1 and LTC2 from lettuce (Bennett et al., 2002), vetispiradiene synthase (VS1) from *Hyoscyamus muticus* (Back and Chappell, 1995), α-humulene synthase (ZSS1) from *Zingiber zerumbet* (Yu et al., 2008), 5-epi-aristolochene synthase (EAS) from *Nicotiana attenuata* (Bohlmann et al., 2002), and a codon optimized δ-guainene synthase (AcC2) from *Aquilaria crassna* (Kumeta and Ito, 2010) were selected for expression in *E. coli* (Fig. 5A). ADS, AaGAS, LTC1, LTC2, AcC2, ZSS1, codon optimized germacrene A synthase Sc1 from *Solidago canadensis* (Prosser et al., 2002), and codon optimized germacrene A synthase IdGAS from *Ixeris dentata* (Kim et al., 2005) were selected for expression in *S. cerevisiae* (Fig. 5B). All of the selected SeTPSs produced significant quantities

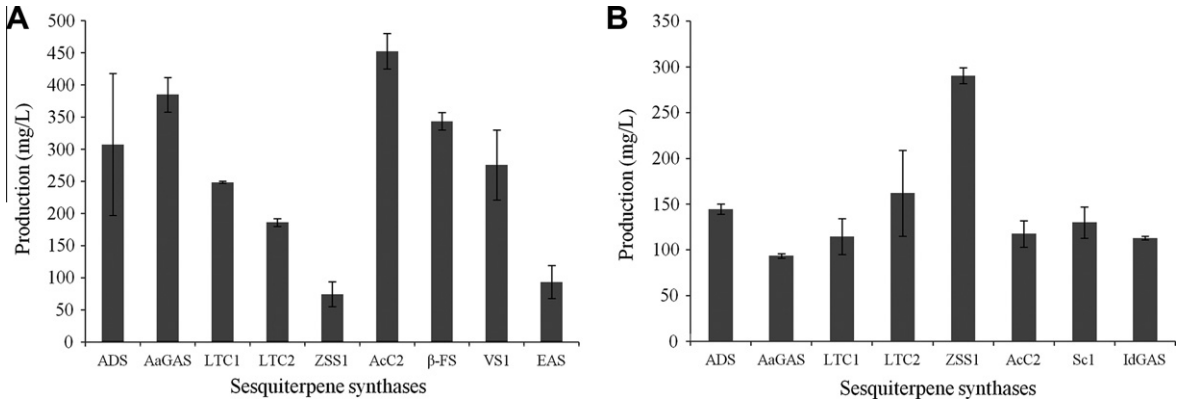

Fig. 5. In vivo sesquiterpene production by expression of different SeTPSs in microbial hosts. (A) Titers of sesquiterpenes produced by *E. coli*, measured at the end of 72-h culture. (B) Titers of sesquiterpenes produced by *S. cerevisiae*, measured at the end of 72-h culture.

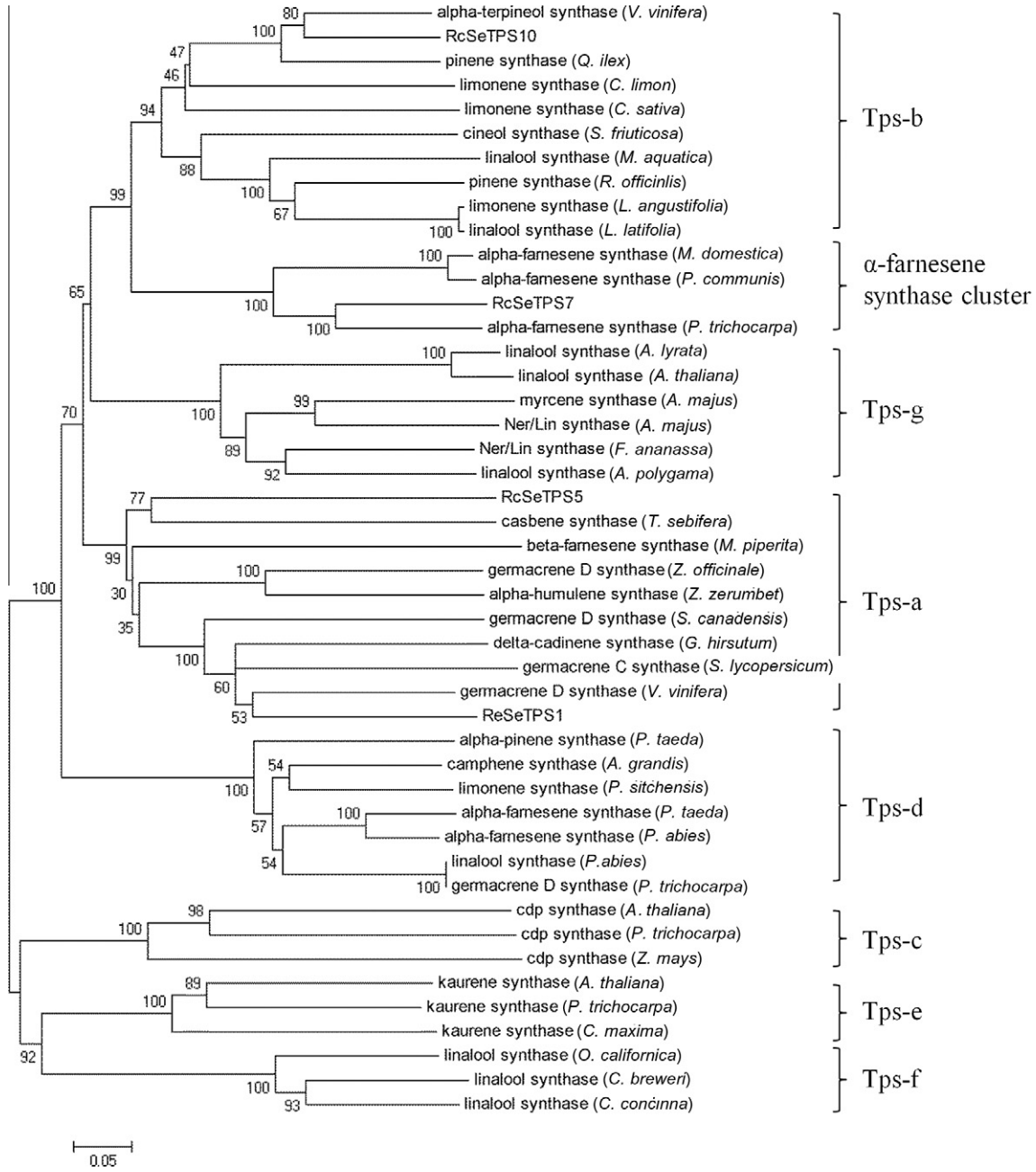

Fig. 6. Phylogenetic tree of RcSeTPS1, RcSeTPS5, RcSeTPS7, RcSeTPS10 together with selected terpene synthases from other plants. The terpene synthases used to build the tree are similar to a selection used previously (Danner et al., 2011). The gene subfamily Tps-a to Tps-g are according to the previous definition (Bohlmann et al., 1998; Dudareva et al., 2003). The alignment was performed by ClustalW algorithm and the tree was built with the neighbor joining method and 1000 replications for bootstrap method. Numbers are actual bootstrap values of branches.

of sesquiterpene products in our hosts, with titers ranging from 75 to 450 mg/L in *E. coli* and 94 to 290 mg/L in *S. cerevisiae*. It was interesting to note that ZSS1 produced the lowest sesquiterpene titer of the SeTPSs expressed in *E. coli*, while it produced the highest titer in yeast (Fig. 5), highlighting the utility of testing both a prokaryotic and eukaryotic host for expression of plant TPS genes. The higher levels of α-humulene observed in yeast may be the result of many factors, including codon usage, protein solubility, mRNA and protein stability, and specific activity in the host environment. The two codon optimized germacrene synthases (Sc1 and IdGAS), however, did not yield higher product titers than the non-codon-optimized germacrene synthases (AaGAS, LTC1, and LTC2), suggesting a limitation in the utility of basic codon optimization (to generate a high codon adaptation index), and consistent with the results obtained with codon optimized RcSeTPS3. Nevertheless, all of the SeTPSs selected from plants other than *R. communis* produced relatively high titers of sesquiterpenes in both hosts, indicating that the SeTPSs isolated from *R. communis* are not as efficient in these optimized hosts for sesquiterpene production.

The cyclization pathway for (−)-α-copaene and (+)-δ-cadinene from RcSeTPS1 is proposed in Fig. 1. To synthesize these two compounds from a single enzyme, the substrate farnesyl pyrophosphate (FPP) (1) could go through ionization, isomerization, and 1,10-cyclization to form the (Z, E)-germacradienyl cation. Because the chirality at C1 is different for (−)-α-copaene (3) (S configuration) and (+)-δ-cadinene (2) (R configuration), the synthetic

pathways may diverge at this point. Route A proceeds through a 1,6-ring closure and deprotonation to produce (+)-δ-cadinene (2). Route B proceeds through a 1,6-ring closure, 2,7-ring closure, 1,3-H^- shift, 1,3-CH_3 shift, 1,2-H^- shift, and deprotonation to produce (−)-α-copaene (3).

2.3. Characterization of the products of RcSeTPSs in vitro

Since only two of the six SeTPSs isolated from *R. communis* produced sesquiterpenes in vivo, the remaining candidates were subjected to in vitro characterization. Following expression in *E. coli BLR* (DE3), RcSeTPS3 and RcSeTPS13 were found to form inclusion bodies, while RcSeTPS5 and RcSeTPS10 were verified to be soluble by SDS-PAGE and purified by Ni-NTA chromatography to 80–90% homogeneity (Fig. S1). The purified proteins were incubated overnight with geranyl pyrophosphate (GPP), FPP (1), or geranylgeranyl pyrophosphate (GGPP), and products were extracted and analyzed by GC-MS. Of the three substrates tested, only incubation of RcSeTPS5 and RcSeTPS10 with FPP yielded terpene products, and in both cases multiple sesquiterpene products were detected (Figs. S2 and S3). RcSeTPS5 produced both sesquiterpene olefins and sesquiterpene alcohols (as judged by the presence of $m/z = 222$ as parent ion, and MS database matches), while RcSeTPS10 produced only sesquiterpene olefins. Due to the lack of standards, and the fact that only one sesquiterpene has been identified from the essential oil of *R. communis*, these products were not identified with any degree of certainty. Instead, the mass spectra for these components and the structures of the best matches from the NIST MS database or the Essential Oils HP–Chemstation library are listed. The lack of sesquiterpene production in vivo by RcSeTPS5 and RcSeTPS10 could be due to insufficient protein levels or slow kinetics, since only low product levels were produced in vitro even in the presence of high protein and substrate concentrations.

2.4. Phylogenetic analysis of RcSeTPSs

The four functional RcSeTPSs, RcSeTPS1, RcSeTPS5, RcSeTPS7, and RcSeTPS10, were subjected to phylogenetic analysis along with several known TPSs from other plants (Fig. 6). RcSeTPS1 and RcTPS5 clustered with the angiosperm SeTPS group Tps-a. The closest homologue of RcSeTPS1 is (−)-germacrene D synthase from *Vitis vinifera* with an identity of 57% (Lucker et al., 2004). The product profiles of these enzymes are very different, but perhaps this is not surprising since they share a very low level of identity. The closest homologue of RcSeTPS5 is casbene synthase from *Triadica sebifera* (*Sapium sebiferum*), a DiTPS with an identity of 43% (Kirby et al., 2010). RcSeTPS7 produced a single acyclic product, (E, E)-α-farnesene (6), and showed the highest in vivo efficiency among all the RcSeTPSs tested. Phylogenetic analysis established that it forms a separate phylogenetic cluster together with other angiosperm α-farnesene synthases from *Malus x domestica*, *Pyrus communis*, and *Populus trichocarpa* (Danner et al., 2011; Pechous and Whitaker, 2004). The new cluster has a closer phylogenetic relationship to the cluster of angiosperm MoTPSs (Tps-b) than the cluster of angiosperm SeTPSs (Tps-a). It is interesting that RcSeTPS10 aligned with the cluster of angiosperm MoTPSs (Tps-b), which was exclusively composed of MoTPSs until two TPSs were demonstrated to have activity against both GPP and FPP (1) (Martin et al., 2010). The difference here is that RcSeTPS10 can only accept FPP (1) as substrate. No plastidial targeting signal peptide was predicted from this protein by the prediction algorithms ChloroP, TargetP, and SignalP. However, not all MoTPSs in the cluster of Tps-b possess signal peptides as evidenced by several functional MoTPSs of *V. vinifera* from this cluster (Martin et al., 2010).

3. Conclusion

In conclusion, cytosolic SeTPSs isolated from *R. communis* using a genome mining strategy were analyzed and functionally characterized. The four functional terpene synthases that were isolated are indeed SeTPSs, producing single, double, and multiple sesquiterpene products. This is the first time that SeTPSs have been characterized from this plant and from the Euphorbiceae as far as we know. In vivo expression of SeTPSs from *R. communis* in our engineered *E. coli* and *S. cerevesiae* hosts yielded significantly lower product titers compared to a selection of other plant SeTPSs. The fact that two RcSeTPSs were only active in vitro may be due to low activity or low protein expression in the hosts or a combination of both.

4. Experimental

4.1. Sequence analysis

The blast algorithms were used to search for SeTPSs in *R. communis* based on sequence similarity in GenBank. Specifically, casbene synthase and neocembrene synthase genes from *R. communis* were used as query sequences to search for TPS genes, which served as additional query sequences to search for more TPS genes in *R. communis*. The predicted cDNAs were used directly for primer design.

4.2. Plant material, mRNA isolation, and cDNA amplification

R. communis was propagated and maintained outdoors as described previously (Kirby et al., 2010). RNA was isolated from untreated plants alongside plants that were treated with the elicitor methyl jasmonate (MeJA). MeJA elicitation was carried out by spraying the whole plant with 100 mL 0.5 mM MeJA in 0.1% Tween-20, and leaf tissue samples were harvested from the plant 20 h after treatment. The MeJA-treated and untreated leaf materials were collected from the plants, immediately frozen in liq. N_2, and ground into a fine power using a mortar and pestle. Total RNA was extracted using RNeasy Plant Mini Kit (Invitrogen). Total RNA (200 pg) was used as a template to synthesize the first strand cDNAs using SuperScript III RT-PCR (Invitrogen). RNaseOUT (Invitrogen) was added to prevent the degradation of mRNA. The single strand cDNAs were used as template for PCR using primers listed in Supplementary Table 1 and subcloned into the Zero pCR-Blunt vector (Invitrogen).

The ADS, VS1, EAS and ZSS1 genes were available in our laboratory. AaGAS and β-FS genes were cloned from an mRNA preparation of *A. annua*. The LTC1 and LTC2 genes were obtained by extracting mRNA from the fresh leaves of lettuce according to the same procedure as RcSeTPS genes. Lettuce was obtained from the University of California, Berkeley Botanical Garden. Sc1, IdGAS, and AcC2 genes were codon optimized for expression in *E. coli* and *S. cerevisiae* and synthesized by GenScript using their optimization algorithm (generating a high codon adaptation index).

4.3. Expression of SeTPSs in S. cerevisiae and product characterization

All RcSeTPS and AaGAS, LTC1, LTC2, Sc1, IdGAS, AcC2, ZSS1 genes were amplified by PCR with introduction of SpeI and SmaI restriction sites. They were first subcloned into the Zero pCR-Blunt vector (invitrogen) for sequence analysis. The correct clones were digested with the above restriction enzymes and ligated into the expression vector pRS426 under the control of the GAL1 promoter (Ro et al., 2006). *S. cerevisiae* EPY300 was used as an expression host for sesquiterpene production (Kirby et al., 2010). Transformation into EPY300 was conducted using the lithium acetate method with selection on synthetic complete agar lacking uracil, histidine, and methionine (Gietz and Schiestl, 2007).

Transformed yeast strains were grown overnight at 30 °C on synthetic complete medium with the appropriate amino acid dropped out. These cultures were inoculated into a modified (0.2% galactose and 1.8% glucose) YPD medium for production with 1 mM L-methionine added to down-regulate squalene synthase (Paradise et al., 2008), overlaid with 10% dodecane, and shaken at 30 °C. At 24, 48, and 72 h, dodecane samples were analyzed by gas-chromatography mass-spectrometry (GC–MS, GC model 6890, MS model 5973 inert, Agilent) to characterize and quantify the sesquiterpene products. The GC analytical method was as follows: hold at 100 °C for 1 min after injection, raise to 250 °C at 50 °C/min, hold at 250 °C for another minute. Because RcSeTPS5 produced multiple products, the GC method was changed at the second step to a gradient of 10 °C/min. Full mass spectra were generated by scanning with the *m*/*z* range from 45 to 230 for sesquiterpene and monoterpene detection, and from 45 to 350 for possible diterpene detection.

4.4. Expression of RcSeTPSs in E. coli DH1 and product characterization

The SeTPS genes were amplified by PCR with introduction of NcoI/BamHI restriction sites for RcSeTPS1, 3, 5, and 7, NheI/BamHI cloning sites for RcSeTPS10 and 14, and NheI/SmaI for AaGAS, LTC1, LTC2, β-FS, VS1, EAS, AcC2, and ZSS1. They were digested with the above restriction enzymes and ligated into pTrc99a for expression. The resulting plasmids were cotransformed with pBbA5c-MevT-MBIS, which contains the complete mevalonate biosynthetic pathway for FPP production, into *E. coli* DH1 (Redding-Johanson et al., 2011).

To produce sesquiterpenes in *E. coli*, the transformants with different SeTPS genes were grown at 37 °C in EZ rich medium (1% glucose) supplied with 30 mg/L chloramphenicol (Cm) and 50 mg/L kanamycin (Kan). At an OD_{600} of 0.5–0.6, expression was induced by adding 1 mM IPTG, and cultures were overlaid with 10% dodecane and moved to a 30 °C incubator for the production of sesquiterpenes. At 24, 48, and 72 h, the samples were checked by GC–MS to characterize and quantify the sesquiterpene products.

4.5. Protein expression from E. coli BLR(DE3) and in vitro enzymatic assays

To purify RcSeTPSs for the in vitro study, RcSeTPS3, 5, 10, and 14 were cloned between the NheI and BamHI sites of the expression vector pSKB3, under control of the T7 promoter. *E. coli* BLR(DE3) transformed with appropriate plasmids was grown in LB medium at 37 °C to an OD_{600} of 0.5, at which time 0.1 mM IPTG was added to the culture and protein expression was carried out at 20 °C for 24 h. Cells were collected by centrifugation, resuspended in Buffer A (50 mM Tris–HCl, pH. 8.0, 2 mM DTT, 2 mM EDTA), and lysed by sonication. Cell debris and insoluble proteins were removed by centrifugation (10,000*g*, 4 °C, 1 h). To the cleared lysate, 2 mL of Ni-NTA resin (Qiagen) were added and SeTPSs were purified using a step gradient of buffer A with increasing concentrations of imidazole. Pure (>90%) proteins were eluted using buffer A containing 250 mM imidazole, buffer exchanged into Buffer A without imidazole, concentrated, aliquoted, and flash frozen. Only RcSeTPS5 and RcSeTPS10 were expressed in a soluble form and purified successfully, while RcSeTPS3 and RcSeTPS14 accumulated as insoluble protein in inclusion bodies.

The purified enzymes were used for in vitro assays. RcSeTPS5 or RcSeTPS10 (5 μM) was incubated with 20 μg/ml GPP in 25 mM Hepes pH 7.2, 100 mM KCl, and 10 mM $MnCl_2$; or incubated with 20 μg/ml FPP in 25 mM Hepes pH 7.5, and 20 mM $MgCl_2$; or incubated with 20 μg/ml GGPP in 25 mM Hepes pH 7.2, 100 mM KCl, 10 mM $MgCl_2$ and 10 μM $MnCl_2$ overnight at room temperature and extracted with equal volume of pentane two times (Lee and Chappell, 2008). The extracts were combined and concentrated to 50 μl for analysis by GC–MS.

Acknowledgements

Funding for this work was provided by NIH RC1 GM090980-01.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.phytochem.2012.02.022.

References

Adams, R.P., 1995. Identification of Essential Oil Components by Gas Chromatography/Mass Spectroscopy. Allured publishing corporation.
Back, K., Chappell, J., 1995. Cloning and bacterial expression of a sesquiterpene cyclase from Hyoscyamus-muticus and its molecular comparison to related terpene cyclases. J. Biol. Chem. 270, 7375–7381.
Bennett, M.H., Mansfield, J.W., Lewis, M.J., Beale, M.H., 2002. Cloning and expression of sesquiterpene synthase genes from lettuce (*Lactuca sativa* L.). Phytochemistry 60, 255–261.
Bertea, C.M., Voster, A., Verstappen, F.W.A., Maffei, M., Beekwilder, J., Bouwmeester, H.J., 2006. Isoprenoid biosynthesis in *Artemisia annua*: Cloning and heterologous expression of a germacrene A synthase from a glandular trichome cDNA library. Arch. Biochem. Biophys. 448, 3–12.
Bohlmann, J., Meyer-Gauen, G., Croteau, R., 1998. Plant terpenoid synthases: Molecular biology and phylogenetic analysis. Proc. Nat. Acad. Sci. USA 95, 4126–4133.
Bohlmann, J., Stauber, E.J., Krock, B., Oldham, N.J., Gershenzon, J., Baldwin, I.T., 2002. Gene expression of 5-epi-aristolochene synthase and formation of capsidiol in roots of *Nicotiana attenuata* and N-sylvestris. Phytochemistry 60, 109–116.
Chan, A.P., Crabtree, J., Zhao, Q., Lorenzi, H., Orvis, J., Puiu, D., Melake-Berhan, A., Jones, K.M., Redman, J., Chen, G., Cahoon, E.B., Gedil, M., Stanke, M., Haas, B.J., Wortman, J.R., Fraser-Liggett, C.M., Ravel, J., Rabinowicz, P.D., 2010. Draft genome sequence of the oilseed species *Ricinus communis*. Nat. Biotechnol. 28, 951–956.
Chen, F., Tholl, D., Bohlmann, J., Pichersky, E., 2011. The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J. 66, 212–229.
Chen, F., Tholl, D., D'Auria, J.C., Farooq, A., Pichersky, E., Gershenzon, J., 2003. Biosynthesis and emission of terpenoid volatiles from Arabidopsis flowers. Plant Cell 15, 481–494.
Danner, H., Boeckler, G.A., Irmisch, S., Yuan, J.S., Chen, F., Gershenzon, J., Unsicker, S.B., Kollner, T.G., 2011. Four terpene synthases produce major compounds of the gypsy moth feeding-induced volatile blend of Populus trichocarpa. Phytochemistry 72, 897–908.
Darmanin, S., Wismayer, P.S., Camilleri Podesta, M.T., Micallef, M.J., Buhagiar, J.A., 2009. An extract from *Ricinus communis* L. leaves possesses cytotoxic properties and induces apoptosis in SK-MEL-28 human melanoma cells. Nat. Prod. Res. 23, 561–571.
Dudareva, N., Martin, D., Kish, C.M., Kolosova, N., Gorenstein, N., Faldt, J., Miller, B., Bohlmann, J., 2003. (E)-β-ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. Plant Cell 15, 1227–1241.
Gietz, R.D., Schiestl, R.H., 2007. Quick and easy yeast transformation using the LiAc/SS carrier DNA/PEG method. Nature Protcol. 2, 35–37.
Kim, M.Y., Chang, Y.J., Bang, M.H., Baek, N.I., Jin, J., Lee, C.H., Kim, S.U., 2005. CDNA isolation and characterization of (+)-germacrene A synthase from *Ixeris dentata form* albiflora Hara. J. Plant Biol. 48, 178–186.
Kirby, J., Keasling, J.D., 2009. Biosynthesis of plant isoprenoids: perspectives for microbial engineering. Annu. Rev. Plant Biol. 60, 335–355.
Kirby, J., Nishimoto, M., Park, J.G., Withers, S.T., Nowroozi, F., Behrendt, D., Rutledge, E.J.G., Fortman, J.L., Johnson, H.E., Anderson, J.V., Keasling, J.D., 2010. Cloning of casbene and neocembrene synthases from Euphorbiaceae plants and expression in *Saccharomyces cerevisiae*. Phytochemistry 71, 1466–1473.
Knight, B., 1979. Ricin–a potent homicidal poison. Br. Med. J. 1, 350–351.
Kumeta, Y., Ito, M., 2010. Characterization of delta-guaiene synthases from cultured cells of aquilaria, responsible for the formation of the sesquiterpenes in agarwood. Plant Physiol. 154, 1998–2007.
Lee, S., Chappell, J., 2008. Biochemical and genomic characterization of terpene synthases in *Magnolia grandiflora*. Plant Physiol. 147, 1017–1033.
Lucker, J., Bowen, P., Bohlmann, J., 2004. Vitis vinifera terpenoid cyclases: functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (−)-germacrene D synthase and expression of mono- and sesquiterpene synthases in grapevine flowers and berries. Phytochemistry 65, 2649–2659.
Martin, D.M., Aubourg, S., Schouwey, M.B., Daviet, L., Schalk, M., Toub, O., Lund, S.T., Bohlmann, J., 2010. Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcdna cloning, and enzyme assays. BMC Plant Biol. 10, 226.

Martin, D.M., Faldt, J., Bohlmann, J., 2004. Functional characterization of nine Norway spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily. Plant Physiol. 135, 1908–1927.

Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D., Keasling, J.D., 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nature Biotechnol. 21, 796–802.

Mau, C.J.D., West, C.A., 1994. Cloning of casbene synthase cDNA – evidence for conserved structural features among terpenoid cyclases in plants. Proc. Nat. Acad. Sci. USA 91, 8497–8501.

Paradise, E.M., Kirby, J., Chan, R., Keasling, J.D., 2008. Redirection of flux through the FPP branch-point in *Saccharomyces cerevisiae* by down-regulating squalene synthase. Biotechnol. Bioeng. 100, 371–378.

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

Picaud, S., Brodelius, M., Brodelius, P.E., 2005. Expression, purification and characterization of recombinant (E)-beta-farnesene synthase from *Artemisia annua*. Phytochemistry 66, 961–967.

Prosser, I., Phillips, A.L., Gittings, S., Lewis, M.J., Hooper, A.M., Pickett, J.A., Beale, M.H., 2002. (+)-(10R)-germacrene A synthase from goldenrod, *Solidago canadensis*; cDNA isolation, bacterial expression and functional analysis. Phytochemistry 60, 691–702.

Redding-Johanson, A.M., Batth, T.S., Chan, R., Krupa, R., Szmidt, H.L., Adams, P.D., Keasling, J.D., Soon Lee, T., Mukhopadhyay, A., Petzold, C.J., 2011. Targeted proteomics for metabolic pathway optimization: application to terpene production. Metab. Eng. 13, 194–203.

Ro, D.K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C., Withers, S.T., Shiba, Y., Sarpong, R., Keasling, J.D., 2006. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440, 940–943.

Sallaud, C., Rontein, D., Onillon, S., Jabes, F., Duffe, P., Giacalone, C., Thoraval, S., Escoffier, C., Herbette, G., Leonhardt, N., Causse, M., Tissier, A., 2009. A novel pathway for sesquiterpene biosynthesis from Z,Z-farnesyl pyrophosphate in the wild tomato *Solanum habrochaites*. Plant Cell 21, 301–317.

Sitton, D., West, C.A., 1975. Casbene – anti-fungal diterpene produced in cell-free-extracts of *Ricinus communis* seedlings. Phytochemistry 14, 1921–1925.

Tan, Q.G., Cai, X.H., Du, Z.Z., Luo, X.D., 2009. Three terpenoids and a tocopherol-related compound from *Ricinus communis*. Helv. Chim. Acta 92, 2762–2768.

Tholl, D., Chen, F., Petri, J., Gershenzon, J., Pichersky, E., 2005. Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from Arabidopsis flowers. Plant J. 42, 757–771.

Williams, D.C., McGarvey, D.J., Katahira, E.J., Croteau, R., 1998. Truncation of limonene synthase preprotein provides a fully active 'Pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. Biochemistry 37, 12213–12220.

Yoshikuni, Y., Martin, V.J., Ferrin, T.E., Keasling, J.D., 2006. Engineering cotton (+)-delta-cadinene synthase to an altered function: germacrene D-4-ol synthase. Chem. Biol. 13, 91–98.

Yu, F.N., Okamto, S., Nakasone, K., Adachi, K., Matsuda, S., Harada, H., Misawa, N., Utsumi, R., 2008. Molecular cloning and functional characterization of alpha-humulene synthase, a possible key enzyme of zerumbone biosynthesis in shampoo ginger (*Zingiber zerumbet* Smith). Planta 227, 1291–1299.