



# The santalene synthase from *Cinnamomum camphora*: Reconstruction of a sesquiterpene synthase from a monoterpene synthase

Alice Di Girolamo<sup>a,1</sup>, Janani Durairaj<sup>b,1</sup>, Adèle van Houwelingen<sup>c</sup>, Francel Verstappen<sup>a</sup>, Dirk Bosch<sup>c</sup>, Katarina Cankar<sup>c</sup>, Harro Bouwmeester<sup>d</sup>, Dick de Ridder<sup>b</sup>, Aalt D.J. van Dijk<sup>b,e</sup>, Jules Beekwilder<sup>c,\*</sup>

<sup>a</sup> Laboratory of Plant Physiology, Department of Plant Sciences, Wageningen University, Netherlands

<sup>b</sup> Bioinformatics Group, Department of Plant Sciences, Wageningen University, Netherlands

<sup>c</sup> Bioscience, Wageningen Plant Research, Netherlands

<sup>d</sup> Swammerdam Institute for Life Sciences, University of Amsterdam, Netherlands

<sup>e</sup> Biometris, Department of Plant Sciences, Wageningen University, Netherlands

## ARTICLE INFO

### Keywords:

Terpene

Monoterpene synthase

Sesquiterpene synthase

Hybrid enzymes

Santalene synthase

*Cinnamomum camphora*

## ABSTRACT

Plant terpene synthases (TPSs) can mediate formation of a large variety of terpenes, and their diversification contributes to the specific chemical profiles of different plant species and chemotypes. Plant genomes often encode a number of related terpene synthases, which can produce very different terpenes. The relationship between TPS sequence and resulting terpene product is not completely understood. In this work we describe two TPSs from the Camphor tree *Cinnamomum camphora* (L.) Presl. One of these, CiCaMS, acts as a monoterpene synthase (monoTPS), and mediates the production of myrcene, while the other, CiCaSSy, acts as a sesquiterpene synthase (sesquiTPS), and catalyses the production of  $\alpha$ -santalene,  $\beta$ -santalene and trans- $\alpha$ -bergamotene. Interestingly, these enzymes share 97% DNA sequence identity and differ only in 22 amino acid residues out of 553. To understand which residues are essential for the catalysis of monoterpenes resp. sesquiterpenes, a number of hybrid synthases were prepared, and supplemented by a set of single-residue variants. These were tested for their ability to produce monoterpenes and sesquiterpenes by *in vivo* production of sesquiterpenes in *E. coli*, and by *in vitro* enzyme assays. This analysis pinpointed three residues in the sequence which could mediate the change in product specificity from a monoterpene synthase to a sesquiterpene synthase. Another set of three residues defined the sesquiterpene product profile, including the ratios between sesquiterpene products.

## 1. Introduction

Plant species have highly specific chemical profiles, which are often determined by the presence of different terpenes and other secondary metabolites. These profiles serve the plant to function in its ecological niche, but have also been widely employed in human applications, including pharmaceuticals, nutraceuticals, food and cosmetics [1]. In some cases, such applications lead to a high demand for the natural source of the metabolites and put increasing pressure on the conservation of the plant species from which they are derived [2–4]. An example

of such an endangered species, which has been over-exploited for the extraction of their essential oils, is the Sandalwood tree. Its heartwood oil is predominantly composed of sesquiterpenes with desirable odour characteristics, and has been used for fragrances and perfumes.

Terpenes, such as found in sandalwood oil, belong to a large class of metabolites and can be distinguished by their number of carbon atoms. Monoterpenes (C<sub>10</sub> compounds) and sesquiterpenes (C<sub>15</sub>) are synthesized by the condensation of two or three isoprene units respectively. Plants produce monoterpenes in plastids, while sesquiterpenes are generally synthesized in the cytosol. These two cellular compartments

**Abbreviations:** TPS, terpene synthase; monoTPS, monoterpene synthase; sesquiTPS, sesquiterpene synthase; IDP, isopentenyl diphosphate; DMADP, dimethylallyl diphosphate; GDP, geranyl diphosphate; FDP, farnesyl diphosphate; K.I, Kovats retention index.

\* Corresponding author. Bioscience, Wageningen Plant Research, Netherlands.

E-mail addresses: [alice1.digirolamo@wur.nl](mailto:alice1.digirolamo@wur.nl) (A. Di Girolamo), [janani.durairaj@wur.nl](mailto:janani.durairaj@wur.nl) (J. Durairaj), [h.j.bouwmeester@uva.nl](mailto:h.j.bouwmeester@uva.nl) (H. Bouwmeester), [dick.deridder@wur.nl](mailto:dick.deridder@wur.nl) (D. de Ridder), [aaltjan.vandijk@wur.nl](mailto:aaltjan.vandijk@wur.nl) (A.D.J. van Dijk), [jules.beekwilder@wur.nl](mailto:jules.beekwilder@wur.nl) (J. Beekwilder).

<sup>1</sup> These authors contributed equally to this work.

<https://doi.org/10.1016/j.abbi.2020.108647>

Received 1 July 2020; Received in revised form 25 September 2020; Accepted 20 October 2020

Available online 26 October 2020

0003-9861/© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

have different pathways responsible for the production of isoprene units. In the cytosol of plants, the mevalonate (MVA) pathway is active, while the 2-C-methyl-D-erythriol-4-phosphate (MEP) pathway operates in plant plastids [5–7]. Both these pathways produce isopentenyl diphosphate (IDP) and dimethylallyl diphosphate (DMADP), which are further condensed into allylic diphosphate substrates such as geranyl diphosphate (GDP) and farnesyl diphosphate (FDP) [8,9]. Subsequently, terpene synthases (TPSs) convert these substrates to monoterpenes or sesquiterpenes, respectively. Precursors for terpenes such as IDP, DMADP, FDP and GDP can be efficiently produced in industrial microorganisms, using similar pathways [10]. Formation of terpenoids from these precursors can also be achieved in microbes, and relies on TPSs, often derived from plants [11,12].

A wide variety of synthases have been described for mono- and sesquiterpenes [6,13,14]. Plant genomes often encode 10–50 different synthases, which determine the terpene profile in the essential oil of the species, and often different chemotypes arise by diversification of terpene synthases. However, it is still challenging to predict the specific terpene produced by a synthase, the roles of the residues present in the catalytic pocket of the synthase and their involvement in determining the product profile. Overall sequence identity is higher between enzymes with different product specificity, belonging to the same species, compared to enzymes from different species producing the same compound. A better understanding of the relationship between individual residues the primary sequence of a TPS and its product profile will facilitate a prediction of uncharacterized synthases [13].

In this study, we set out to isolate a santalene synthase. Sandalwood oil has high value for perfumery and is traditionally extracted by steam distillation of *Santalum album* trees older than 15 years [15]. *S. album* has been listed as a vulnerable species in the IUCN Red list of Threatened species [16] and the use of this tree for sandalwood oil extraction has been strictly regulated. The four main compounds present in sandalwood oil are  $\alpha$ -,  $\beta$ -, and epi- $\beta$ -santalol and  $\alpha$ -bergamotol, which are the hydroxylated analogues of  $\alpha$ -,  $\beta$ -, and epi- $\beta$ -santalene and  $\alpha$ -bergamotene respectively. Hydroxylation of these compounds to their alcohols is mediated by cytochrome P450s [17]. The TPSs responsible for the production of santalenes have been identified in two *Santalum* species: *S. album* and in *S. spicatum* [18–20]. In the current work we isolate a santalene synthase from a completely unrelated tree, *Cinnamomum camphora*. The essential oil of one of *C. camphora* chemotypes has been observed to contain santalenes. One of the chemotypes of this species also produces santalenes in its essential oil [39]. The gene encoding the santalene synthase was isolated together with a highly related gene encoding a monoterpene synthase. These two enzymes, displaying different substrate and product specificity despite their high sequence identity, were used to study the role of individual residues in determining substrate and product specificity.

We demonstrate how few residue positions are responsible for substrate specificity, allowing a monoTPS to acquire sesquiTPS activity, without losing its original function. With the results, new insights on functional residues were obtained, contributing to the larger framework of TPSs substrate and product specificity prediction.

## 2. Methods

### 2.1. Identification of santalenes in *C. camphora*

A *C. camphora* plant of the cineole type was purchased from Planfor (France). Leaves, stems and roots were dissected, and 0.5 g of plant material was weighted in a pre-cooled glass tube and suspended in 2 mL dichloromethane. The samples were vortexed for 1 min, sonicated for 5 min in an ultrasonic bath and centrifuged at 1500 g at room temperature to separate the plant material from the supernatant. 1 g Na<sub>2</sub>SO<sub>4</sub> columns were used to dry the obtained supernatant. About 2  $\mu$ L was analysed by gas chromatography mass spectrometry (GC-MS) as previously described [21,22]. All compounds were identified using the mass spectra

deposited in the NIST library and confirmed using their retention index, or by comparison to an original standard, when available (Fig. 1). Santalenes were further confirmed by comparison of retention times and mass spectra of a sandalwood oil standard (Merck, Germany).

### 2.2. RNA extraction from root tissue

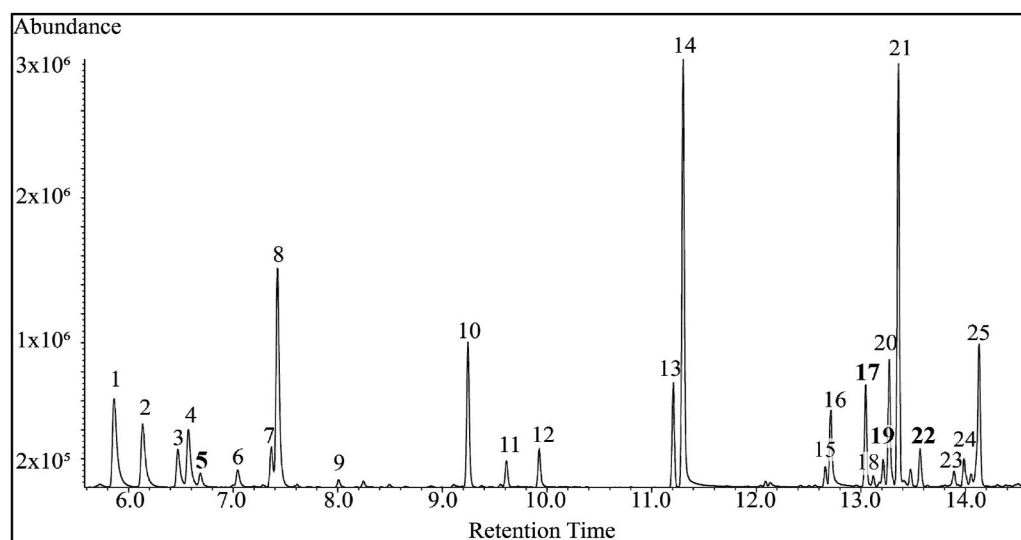
To extract RNA from the root material of *C. camphora*, an extraction buffer was prepared (2% hexadecyl-trimethylammonium bromide, 2% polyvinylpyrrolidone K 30, 100 mM Tris-HCl (pH 8.0), 25 mM EDTA, 2.0 M NaCl, 0.5 g/L spermidine and 2%  $\beta$ -mercaptoethanol). 3 g of ground tissue was mixed to 15 mL of pre-warmed (65 °C) extraction buffer. The mixture was extracted twice with an equal volume of chloroform:isoamylalcohol (1:24), and ¼ volume of 10 M LiCl was mixed to the supernatant. The RNA was precipitated overnight at 4 °C and harvested by centrifugation at 10000g for 20 min. The pellet was dissolved in 500  $\mu$ L of STE buffer [1.0 M NaCl, 0.5% SDS, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (pH 8.0)] and extracted once with an equal volume of chloroform: isoamylalcohol. Two volumes of ethanol were added to the supernatant, incubated for at least 2 h at –20 °C, centrifuged at 13000 rpm and the supernatant removed. The pellet was air-dried and resuspended in water. Total RNA (60  $\mu$ g) was shipped to Vertis Biotechnology AG (Freising, Germany). PolyA + RNA was isolated, and cDNA was synthesized using a randomized N6 adapter primer and M-MLV H-reverse transcriptase (Sigma). TPS sequences were identified by a cDNA sequencing approach, as described in detail by Beekwilder et al. (2014) [14].

### 2.3. Isolation of CiCaMS and CiCaSSy

Full length open reading frames of putative TPSs were amplified from the cDNA of *C. camphora*. Specific primers (CCH\_TS23\_fw and CCH\_TS23\_re, see Table S1) were designed to amplify total ORFs with a 6-His tag fused at the N-terminus in the plasmid pCDF-Duet1 (Novagen, Merck Chemicals B.V., Amsterdam, the Netherlands). Two variants, namely pCDF-CiCaMS or pCDF-CiCaSSy, were cloned using the same primer pair using BamHI and NotI restriction enzymes. Amplification of cDNA ends (5' RACE; Clontech cDNA RACE kit) experiments were performed to isolate longer versions of both genes, but no cDNA ends with more upstream start codons could be identified. Sequences were deposited in Genbank under accession numbers MN756611 (CiCaMS) and MN756612 (CiCaSSy).

### 2.4. Cloning of CiCaMS/CiCaSSy hybrids and single mutants

To obtain the hybrid proteins of the parental enzymes CiCaMS and CiCaSSy, a library of fragments was designed. When possible, separate fragments were amplified using pCDF-CiCaMS or pCDF-CiCaSSy as template (minimum fragment length was 150bp). Each region was designed to contain 3–6 amino acid substitutions. In Fig. 3A a schematic representation of the design is shown. Table S2 reports the fragment composition of each hybrid, while the primer used for amplification and sequencing are listed in Table S1A. All primers were supplied by IDT (Leuven, Belgium). For fragment amplification, including the vector backbone (derived from Novagen commercial plasmid pACYC-Duet1), Q5 High Fidelity polymerase by NEB was used, following the protocol provided by the supplier. For hybrids assembly, the Circular Polymerase Extension Cloning (CPEC) method was used [23]. For fragments smaller than 150bp and for single mutants, the QuickChange site-directed mutagenesis protocol described by Xia et al. [24] was adapted for the use of Q5 HiFi polymerase, as described in the NEB protocol (Q5 site-directed mutagenesis kit protocol, E0554). All single mutants are listed in Table S3, including the primers used. Before transformation in DH5a, the products obtained from QuickChange were digested with DpNI to eliminate the traces of template. All constructs were confirmed by sequencing from MacroGen. All plasmids used in this study are listed



**Fig. 1.** GC-MS analysis of pentane extract of *Cinnamomum camphora* root tissue. The compounds were identified by their mass spectra using the NIST library and confirmed comparing their retention indexes with the reference list provided by Adams [44]. The y-axis reports the GC-MS response units, while the x-axis reports the retention times. The compounds identified for each peak are listed. 1.  $\alpha$ -pinene; 2. camphene; 3. sabinene; 4.  $\beta$ -pinene; 5. **myrcene**; 6.  $\alpha$ -pinene; 7. limonene; 8. cineole; 9. cis- $\beta$ -terpineol; 10. camphor; 11. borneol; 12.  $\alpha$ -terpineol; 13. bornyl acetate; 14. safrene; 5.  $\alpha$ -cubebene; 16.  $\beta$ -elemene; 17.  **$\alpha$ -santalene**; 18. trans- $\beta$ -caryophyllene; 19. **trans- $\alpha$ -bergamotene**; 20.  $\alpha$ -guaiane; 21. guaia-6,9-diene; 22.  **$\beta$ -santalene**; 23. germacrene D; 24.  $\beta$ -selinene; 25.  $\alpha$ -bulnesene.

in Table S1B.

## 2.5. Heterologous expression of CiCaMS and CiCaSSy in *E. coli* BL21DE3

To analyse the sesquiterpene product profile of the enzymes, an *E. coli* expression strain BL21DE3 containing an additional plasmid expressing all genes necessary for the synthesis of FDP (pBbA5c-MevT-MBIS-NPtlI) was used. This plasmid is a variant of plasmid pBbA5c-MevT(CO)-MBIS(CO, IspA) [10,25] in which the chloramphenicol resistance marker has been exchanged for a kanamycin resistance marker (NptII). Another variant, with a different origin of replication (colE1) was also used in the experiments (pBbE5k-MevT(CO)-MBIS(CO)). Fermentations were performed using 20 mL of 2xYT medium (16 g/L tryptone, 10 g/L NaCl, 10 g/L yeast extract) in 100 mL glass flasks. Overnight cultures were diluted to an OD<sub>600</sub> of 0.150 and incubated at 37 °C 250 rpm until an OD<sub>600</sub> of 0.4–0.5 was reached, then IPTG 1 mM and 2 mL dodecane were added, followed by 24 h incubation at 28 °C 250 rpm. A concentration of 50  $\mu$ g/mL kanamycin and 50  $\mu$ g/mL chloramphenicol was used to maintain the plasmids in the system. The 2 mL dodecane was then recovered for GCMS analysis by centrifugation at 3600 rpm for 15 min. For the GC-MS analysis, 20–80 mg dodecane were weighted and diluted in 2 mL ethyl acetate. This solution was dried over a Na<sub>2</sub>SO<sub>4</sub> column before analysis.

To confirm the results obtained with the fermentation analysis and to assess the monoterpene activity of the enzymes, *in vitro* enzyme assays were performed. The BL21DE3 *E. coli* expression strain was used for protein production. Overnight cultures were diluted to an OD<sub>600</sub> of 0.150 in 20 mL 2xYT and incubated for at 37 °C 250 rpm until an OD<sub>600</sub> of 0.6–0.8 was reached. A concentration of 1 mM IPTG was added and the cultures were grown at 18 °C 250 rpm overnight. Cells were then harvested by centrifugation (10 min, 3600 rpm), medium was removed, and cells were resuspended in 1 mL Resuspension buffer (50 mM Tris-HCl pH = 7.5, 1.4 mM  $\beta$ -mercaptoethanol; 4 °C). Cells were disrupted by shaking 2 times for 10 s with 0.2 g zirconium sand in a Fastprep machine at speed 6.5. Insoluble particles were subsequently removed by centrifugation (10 min 13000 rpm, 4 °C). Soluble protein was immediately used for enzyme assays or stored in a 12.5% glycerol solution.

For enzyme assays, both farnesyl diphosphate and geranyl diphosphate (10 mM, Sigma FDP dry-evaporated and dissolved in 50% ethanol) were used as substrates. In a glass tube, a mix was made of 800  $\mu$ L of MOPSO buffer (15 mM MOPSO (3-[N-morpholino]-2-hydroxypropane sulphonic acid) pH = 7.0, 12.5% glycerol, 1 mM MgCl<sub>2</sub>, 0.1% tween 20, 1 mM ascorbic acid, 1 mM dithiothreitol). 100  $\mu$ L of crude enzyme extract and 5  $\mu$ L of FDP or GDP and 20  $\mu$ L Na-orthovanadate

250 mM. 1 mL pentane was added to the mix to extract the terpenes. This mix was incubated at 30 °C with mild agitation for 2–4 h. Subsequently, the mix was centrifuged at 1200 g to recover the pentane, which was dried over a Na<sub>2</sub>SO<sub>4</sub> column and analysed by GC-MS.

## 2.6. GCMS analysis

The GC-MS analysis was performed on an Agilent Technologies system, comprising a 7980A GC system, a 597C inert MSD detector (70 eV), a 7683 auto-sampler and injector and a Phenomenex Zebron ZB-5ms column of 30 m length x 0.25 mm internal diameter and 0.25  $\mu$ m stationary phase, with a Guardian precolumn (5 m). In this system, 1  $\mu$ L of the sample was injected. The injection chamber was at 250 °C, the injection was splitless, and the ZB5 column was maintained at 45 °C for 2 min after which a gradient of 10 °C per minute was started, until 300 °C. Peaks were detected in chromatograms of the total ion count. Compounds were identified by their retention index and by their mass spectrum in combination with comparison of the mass spectrum to libraries (NIST and in-house). The data obtained with the GCMS analysis were processed in order to obtain the average relative concentration of products produced, with the summed area of all peaks representing 100%, and the function “st.dev.s.” of Excel was used to calculate the standard deviation. For concentrations, dodecane samples from three independent cultures were diluted 1:10 in acetone, and analysed by GCMS, using split 10 injection. Concentrations were calculated by comparison of peak areas of selected ions (*m/z* 69, 93, 94, 119, 122, 204) in samples and in standard curves of santalene oil ingredients (kindly provided by Celina Vossen). For total sesquiterpene concentrations, values for  $\alpha$  santalene,  $\beta$  santalene and trans  $\alpha$  bergamotene were added.

## 2.7. Expression of CiCaMS and CiCaSSy in plants

Transient expression in *Nicotiana benthamiana* was employed to determine the product spectrum of the enzymes in a plant heterologous system. CiCaSSy and CiCaMS coding regions were amplified from the pACYC-constructs, using primers CCHattB1-FW and CCHattB2-RE (Table S1). The genes were then cloned using Gateway technique [26], into pBINplus [27]. pBINplus was taken along as a negative control. The obtained plasmids were confirmed by sequencing and transformed to *Agrobacterium tumefaciens* AGL0 via electroporation.

For agroinfiltration, the transformed *A. tumefaciens* AGL0 cultures were grown overnight in LB medium at 28 °C, 300 rpm. 200  $\mu$ L of O/N culture was transferred in 10 mL LB + 10 mM MES + 40  $\mu$ M

acetosyringone and grew again for 16 h 28 °C, 300 rpm. The cultures were then centrifuged at 4000 rpm for 10 min, and the pellets were resuspended in a 10 mM MgCl<sub>2</sub> solution, at a final OD<sub>600</sub> of 1.5. Acetosyringone was added at a final concentration of 200 µM. The suspension was left at room temperature with no shaking for at least 3 h before performing the agroinfiltration.

Young leaves from 3 to 4 weeks old *N. benthamiana* were selected for agroinfiltration. The experiment was performed using biological and technical replicates. Each leaf was infiltrated with about 1 mL *A. tumefaciens* suspension.

Trapping of headspace volatiles was performed as described [22] with following modifications: headspace sampling was performed in a climate room (20 ± 2 °C, 56% RH) with LED lighting (adjusted at 100% white, 10% deep red, 100% far red and 5% blue light). Volatiles were trapped by sucking air out of the jar at a rate of 100 mL/min (inlet flow at 150 mL min<sup>-1</sup>) for 4 h.

Trapped headspace volatiles were analysed using a Thermo TraceGC Ultra connected to a Thermo TraceDSQ quadrupole mass spectrometer (Thermo Fisher Scientific, Waltham, USA). Settings as described [21], with the following modification: volatiles were injected on the analytical column at split ratio 300. Products were identified using original standards (myrcene standard and sandalwood oil, Sigma, Amsterdam, Netherlands), according to their retention time and mass spectra.

## 2.8. SDS-PAGE gel preparation

SDS-Polyacrylamide Gel Electrophoresis was used for protein visualization. The protocol was adapted from Sambrook and Russell, (2006) [28]. 10% resolving SDS gels were prepared using 4 mL water, 3 mL 30% acrylamide mix (Bio-Rad), 2.5 mL Tris-HCl (1.5 M, pH 8.8), 0.1 mL 10% SDS, 0.1 10% ammonium persulfate and 4 µL TEMED (Bio-Rad) for a total of 10 mL solution. Stacking gels were prepared as 5% SDS, using 0.68 mL water, 0.17 mL 30% acrylamide mix, 0.13 mL Tris-HCl (1 M, pH 6.8), 10 µL 10% SDS, 10 µL 10% ammonium persulfate and 1 µL TEMED (Bio-Rad) for a total of 1 mL solution. The protein content in the crude extracts were quantified using Bradford reagent and loaded in equal amount in 5X SDS-gel loading buffer. Precision Plus Protein (Bio-Rad) marker was used and the gels were stained in a 30% ethanol, 8% acetic acid and 0.05% (w/v) coomassie brilliant blue for 2–6 h. Destaining was performed in demineralized water overnight.

## 2.9. Modelling of a 3D structure for CiCaSSy

A homology model of CiCaSSy was created using multi-template modelling. The templates used were the *Hyoscyamus muticus* pre-naspirodiene synthase (PDB ID: 5JO7), *Mentha spicata* limonene synthase (PDB ID: 2ONH), and *Citrus sinensis* limonene synthase (PDB ID: 5UV2); these were selected based on their high sequence similarity with the two *Cinnamomum* synthases. MODELLER [29] was used to create 500 models using the default automodel approach, and the model with the best N-DOPE score was chosen for further analysis. Furthermore, the position of an analogue of FDP, trifluorofarnesyl diphosphate (FFF) in the model was obtained by superposing the 5-epi-aristolochene synthase from *Nicotiana tabacum* (PDB ID: 5EAU), using the align command of PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. Residues differing between the two *Cinnamomum* synthases were visualized using PyMOL. To address the quality of the model, the same command was also used to align the crystal structure of *Santalum album* santalene synthase (SaSSy, PDB ID: 5ZZJ) to the modelled CiCaSSy. The structure alignment of CiCaSSy with SaSSy reveals a high structure similarity between the two enzymes, despite their low sequence identity. The CiCaSSy model can be superposed to SaSSy with a root-mean-square deviation (RMSD) of 1.09 Å over 420 residues.

## 2.10. Phylogenetic tree construction

A phylogenetic tree was constructed to identify the TPS subfamily in which CiCaMS and CiCaSSy belong. 268 sesquiTPSs from our previously assembled database [13], along with 59 monoTPSs, and 27 diTPSs from SwissPROT [30], were clustered into groups of up to 70% sequence identity using CD-HIT [31]. Representative sequences were taken from each cluster and aligned along with CiCaMS and CiCaSSy using Clustal Omega [32] with the Pfam [33] domains Terpene synth (Pfam ID: PF01397) and Terpene synth\_C (Pfam ID: PF03936) as guides for the alignment. The alignment was preprocessed with trimAl [34] such that columns with over 50% gaps were discarded. The tree was constructed using the ETE3 python library [35] with the pmodeltest-ultrafast and RAXML [36] options, and visualized using iTOL [37]. The TPS subfamily assignment was done as in Chen et al. [38]. The phylogenetic tree is represented in Fig. S4.

## 3. Results

### 3.1. Isolation and characterization of CiCaMS and CiCaSSy

The cineole chemotype of *C. camphora* has been observed to contain santalenes [39,40], while other chemotypes (camphor, linalool) have not been reported to contain santalenes [41–43]. Extracts from different parts of a plant from the cineole chemotype were analysed by GC-MS. Roots, leaves and stem of *C. camphora* appeared to contain compounds that correspond to α-santalene, trans-α-bergamotene, and β-santalene, among other compounds, as shown in Fig. 1. The concentration of santalenes was highest in the roots, therefore this tissue was further selected for RNA extraction and cDNA sequencing.

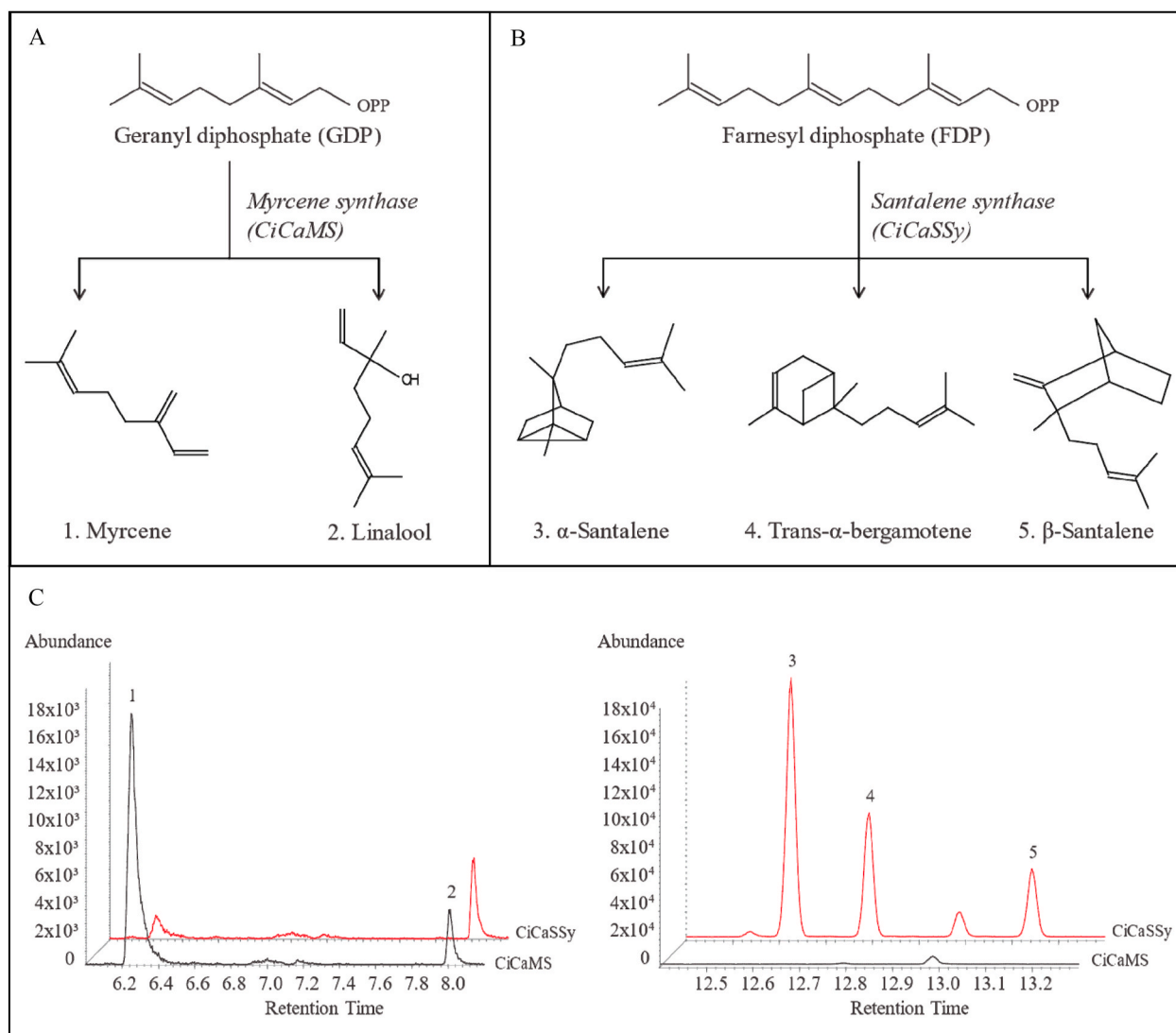
Among the root cDNAs found to correspond to TPSs, one sequence was identified as a putative santalene synthase. When this synthase-encoding sequence was amplified from root cDNA, two different sequences were cloned, which were 97% identical at the DNA level, and encoded proteins which differed in 22 out of 553 amino acids (95% identity).

The enzyme activity of both variants was investigated *in vitro*, by testing product formation using GDP or FDP as substrate. One variant converted FDP to santalenes and was referred to as *Cinnamomum camphora* Santalene Synthase (CiCaSSy). Products were identified as α-santalene, trans-α-bergamotene, and β-santalene, by their mass spectra and by comparison to a sandalwood oil standard. The other variant produced a monoterpene when GDP was used as a substrate and was referred to as CiCaMS. The product of CiCaMS was identified as myrcene, by comparison to a myrcene standard and using the retention time and the retention index [44] (Fig. 2).

The product profile of the two synthases was confirmed by over-expression in a plant system. For expression in *N. benthamiana*, both CiCaSSy and CiCaMS full length coding regions were cloned into binary vectors and the effect of their transient expression on the headspace of *N. benthamiana* was investigated (Fig. S1). Expression of CiCaSSy led to the presence of α-santalene, trans-α-bergamotene, and β-santalene in the headspace, while no monoterpene formation was observed (Fig. S2E). Transient expression of CiCaMS led to emission of myrcene (Fig. S2C), while no sesquiterpene formation could be observed. Standards were used for identification of the compounds (Fig. S2B/D). Therefore, CiCaSSy was identified as a santalene synthase, while CiCaMS was identified as a myrcene synthase.

A BLAST analysis in the non-redundant database [45] revealed that CiCaSSy was very closely related (92% identical) to a predicted (–)-α-terpineol synthase from *Cinnamomum micranthum* [46]. The closest characterized synthases include an (–)-α-terpineol synthase from *Magnolia grandiflora* [47], with 46% identity, and an (–)-α-terpineol synthases from *Vitis vinifera* [48] with 43%. Sequence identity to sesquiTPSs such as *Santalum album* and *Santalum spicatum* santalene synthase (SaSSy and SspSSy) was low (37.8% and 38%; [20]) (Fig. S3).





**Fig. 2.** (A) Conversion of GDP into 1. myrcene and 2. linalool by CiCaMS and (B) conversion of FDP into 3.  $\alpha$ -santalene, 4. trans- $\alpha$ -bergamotene and 5.  $\beta$ -santalene by CiCaSSy. (C) Chromatograms of *in vitro* enzyme assay GC-MS analysis. The graphic on the left shows the production using GDP as substrate for CiCaMS and CiCaSSy. Products were identified as 1. myrcene (K.I. 991) and 2. linalool (K.I. 1098). The graphic on the right shows the production using FDP as substrate for CiCaMS and CiCaSSy. No sesquiterpene activity was detected for CiCaMS, while products of CiCaSSy were identified as 3.  $\alpha$ -santalene (K.I. 1420); 4. trans- $\alpha$ -bergamotene (K.I. 1436) and 5.  $\beta$ -santalene (K.I. 1462). The chromatograms are obtained by extracting ion 93 from the total ion count.

Thus, it appeared that CiCaSSy was most closely related to monoTPSs, rather than sesquiTPSs. An alignment of 175 synthases was generated to construct a phylogenetic tree (Fig. S4). By following the clades proposed by Chen [38] we identified CiCaMS and CiCaSSy to belong to the TPS-g clade.

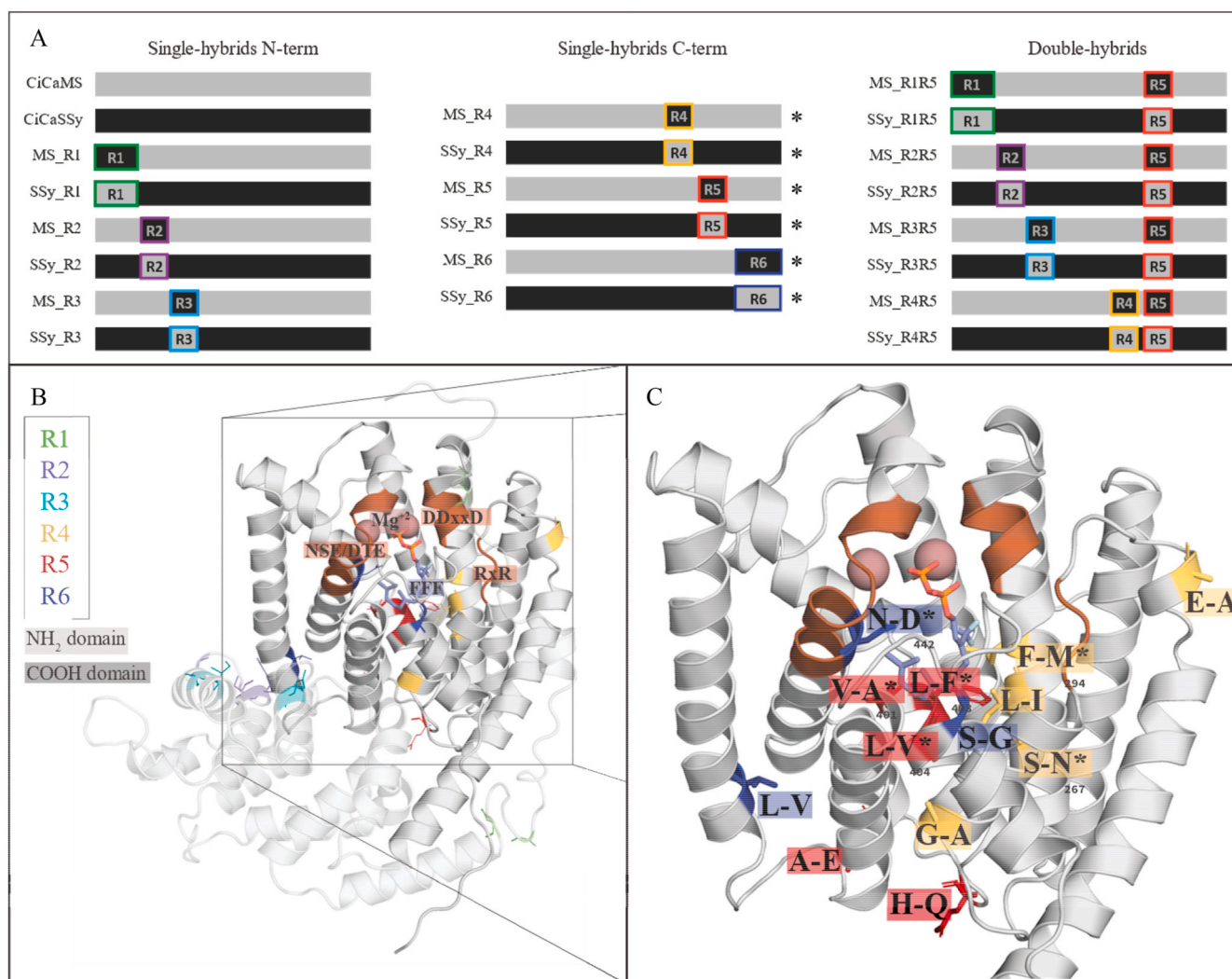
### 3.2. Activity of CiCaMS/CiCaSSy hybrids for production of sesquiterpenes in *E. coli*

Mono- and sesquiTPSs in plants have a common structural fold. They consist of two domains, an N-terminal domain, which is not part of the catalytic site, and a C-terminal domain forming the hydrophobic active site cavity [12]. The C-terminal domain region contains two  $Mg^{2+}$  coordination sites at its opening (labelled DDxxD and NSE/DTE in Fig. 3B) and a loop containing the diphosphate binding site for the substrate (labelled RxR in Fig. 3B). A structure model for CiCaSSy was generated using existing TPSs crystal structures as templates. Fig. 3 highlights the regions which display variation between CiCaSSy and CiCaMS. The genes encoding CiCaMS and CiCaSSy were divided into six regions based on the position of the substitutions in the linear protein sequence.

Regions R1, R2 and R3 are located in the N-terminal domain of the protein. In our 3D model of CiCaSSy (Fig. 3B), these three regions are coloured green, purple, and light blue respectively. The residue substitutions in these regions are distant from the substrate binding cavity, and are therefore unlikely to affect product formation. Regions R4 and R5, and to a lesser extent R6 are close to the active site.

This information was used to investigate which regions of CiCaSSy and CiCaMS play a role in substrate- and/or product specificity. Hybrids exchanging one or two regions between both enzymes were generated (Fig. 3A). Initially, hybrids were tested by *in vitro* enzyme assays, using cell-free extracts, and FDP or GDP as a substrate (Figs. S6A–B, Fig. 2C). Wild type CiCaMS hardly produces sesquiterpenes when FDP is supplied as a substrate. CiCaSSy produces some linalool when GDP is supplied as a substrate. Interestingly, we did not observe any linalool formation by CiCaSSy when expressed in *N. benthamiana* (Figure S1 B–C).

As a first step, substrate specificity of hybrids was investigated, by determining the ratio between monoterpene and sesquiterpene products (Fig. 4). While exchanging regions R1, R2, R3 or R6 did not have a strong effect on the ratio between monoterpenes and sesquiterpenes produced, exchange of region R4 or R5 appeared to have a profound effect on both



**Fig. 3.** (A) Schematic representation of hybrid design. The monoTPS (CiCaMS) is represented in gray, while the sesquiTPS is represented in black. The region swaps are highlighted and depicted as for the parental enzyme (from R1 to R6, respectively: green, purple, light blue, orange, red, ocean blue). The asterisks indicate the regions where single residue exchanges were made. (B) 3D model of CiCaSSy, including Mg<sup>2+</sup> ions [pink] and FDP analogue (trifluorofarnesyl diphosphate, FFF) [sticks]. The model was generated with Modeller [29], using three templates, of which: one sesquiTPS (5JO7) and two monoTPSs (2ONH and 5UV2). The different regions are highlighted in different colors as described in (A). (C) Zoom in into the active site of CiCaSSy. R4, R5 and R6 are visible. In the picture, all amino acid substitutions of the three regions are named, from CiCaMS to CiCaSSy. The asterisks point out the residues essential for the evolution of CiCaMS into a santalene synthase. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

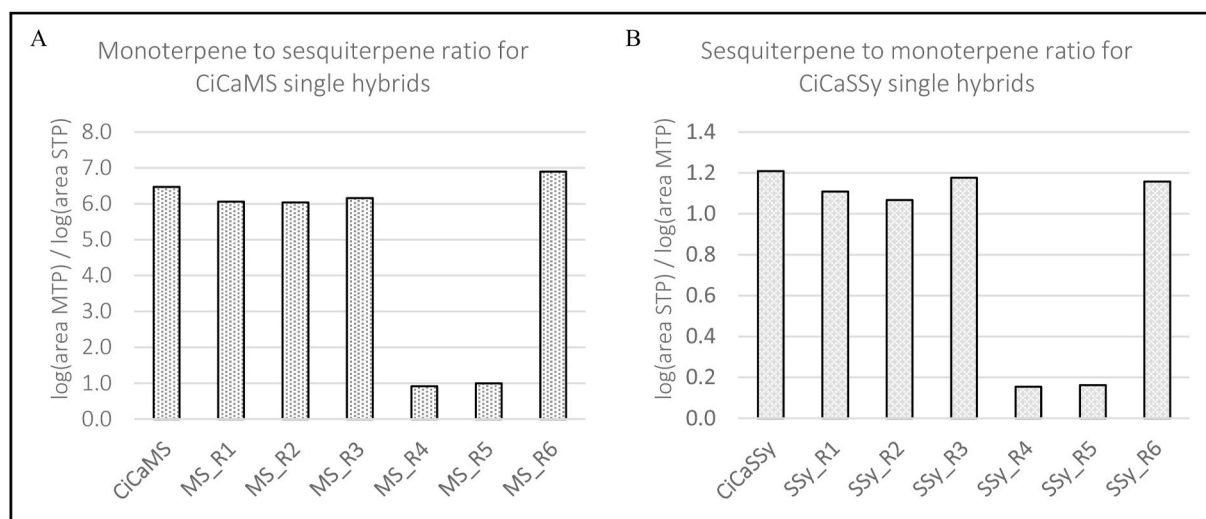
sesquiterpene and monoterpene synthase, as it completely reversed the substrate specificity of the enzyme: MS\_R4 and MS\_R5 use FDP as a substrate, leading to sesquiterpenes production, while SSy\_R4 and SSy\_R5 do not use FDP. Apart from having a strong effect on substrate preference (as observed for R4 and R5 in Fig. 4), exchange of these regions also had an effect on the product profile.

Monoterpene profiles of hybrid enzymes were derived from *in vitro* assays supplying GDP as a substrate (Fig. 5A–B). All CiCaSSy-derived hybrids behaved like CiCaSSy, and showed linalool formation, with traces of myrcene. While quantification based on enzyme assays was difficult, in general linalool peaks in CiCaSSy-derived hybrids were very low, compared to myrcene production in CiCaMS-derived hybrids (see Fig. 2C). All CiCaMS-derived hybrids showed a predominance of myrcene production, as reported for CiCaMS. An exception was observed when both region R4 and R5 were exchanged: MS\_R4R5 and SSy\_R4R5 showed product profiles more similar to CiCaSSy and CiCaMS respectively (Fig. 5A–B). Apparently, these regions define the identity of the formed monoterpene.

To exclude that the absence of sesquiTPS activity of some of the

hybrids (e.g. SSy\_R5) was related to poor solubility of the hybrid proteins, an SDS PAGE protein gel analysis was performed of the cell free extracts. This did not reveal obvious differences in the amount of soluble TPS protein between active and inactive synthases (Fig. S8). This supports the hypothesis that hybrids were not compromised in their overall protein folding, and that the lack of sesquiterpene production in some of the hybrids is due to the changes in the structure of the active site of the enzyme.

The sesquiterpene profile of different hybrids was more diverse, and was eventually addressed using an *in vivo* production method. To this end, the WT parent enzymes and the hybrids were expressed in *E. coli*, in combination with a plasmid which supplies FDP, and their performance was tested in a flask fermentation, using a dodecane overlay for collecting product. After fermentation, the dodecane layer was analysed by GC-MS, and the sesquiterpene profile was extracted. In the *in vivo* system, all hybrids displayed sesquiterpene product profiles similar to those observed in the *in vitro* experiments (Figs. 5–7; Figs. S6A–B). No change in sesquiterpene profile was observed for hybrids covering regions R1, R2 and R3. Hybrids covering regions R4, R5 and R6 displayed marked



**Fig. 4.** Calculated ratio for (A) monoterpene vs. sesquiterpene production in CiCaMS and CiCaMS-derived single hybrids and (B) sesquiterpene vs. monoterpene production in CiCaSSy and CiCaSSy-derived single-hybrids. Data from *in vitro* enzyme assays were used. Each enzyme was tested in duplicate. For the ratios, the sum of the areas of the principal compounds was calculated and converted in logarithmic scale. Note that, as all enzymes showed monoterpene activity *in vitro*, graphic (A) displays an overall higher ratio compared to graphic (B).

changes in sesquiterpene profiles.

Region R4 localizes in the C-terminal domain (yellow in Fig. 3C). It contains three substitutions located in the active site and two in its proximity, suggesting that its exchange could have an impact on the resultant terpene profiles. When the R4 region of CiCaSSy was introduced in monoTPS CiCaMS, the hybrid protein (MS\_R4) displayed sesquiTPS activity, producing predominantly trans- $\alpha$ -bergamotene, with some  $\beta$ -santalene (Fig. 5C). Conversely, when the R4 region of CiCaMS was introduced into CiCaSSy (SSy\_R4), all sesquiterpene production was lost (Fig. 5D).

Region R5 (represented in red in Fig. 3C) is only 12 amino acids long but contains five substitutions, three of which are located very close to the substrate in the 3D model. Again, the exchange of R5 caused the complete loss of sesquiterpene synthase activity in CiCaSSy (SSy\_R5, Fig. 5D). On the other hand, introducing region R5 from CiCaSSy into CiCaMS (MS\_R5), resulted in production of trans- $\alpha$ -bergamotene, as was also observed for MS\_R4 (Fig. 5C).

Region R6 (represented in blue in Fig. 3C) is located at the C-terminus of the protein. It comprises three substitutions between CiCaMS and CiCaSSy. Replacing region R6 from CiCaSSy by R6 from CiCaMS (SSy\_R6) resulted in the production of all wild type sesquiterpene products, but with a relatively higher production of trans- $\alpha$ -bergamotene compared to CiCaSSy (Fig. 5D). Introduction of R6 from CiCaSSy into CiCaMS did not result in any sesquiterpene production (MS\_R6) (Fig. 5C).

A set of double hybrids was generated for CiCaMS and CiCaSSy (Fig. 3A), in which regions R1 to R4 were simultaneously exchanged in combination with R5. All double hybrids carrying R1-R3 regions in combination to R5 showed the same product profile as R5 single hybrids (Fig. 56C), indicating that none of these regions contribute to product specificity. However, the double hybrid carrying R4 and R5 from CiCaSSy in the mainframe of CiCaMS (MS\_R4R5) restored the production of all three main products of CiCaSSy. The relative peak ratio of the products of MS\_R4R5 was comparable to the profile produced by SSy\_R6, confirming that the residues essential for restoring CiCaSSy product profiles in CiCaMS are located in R4, R5 and R6 (Fig. 5C–D).

### 3.3. Single-residue mutants identify critical residues for sesquiterpene synthase activity

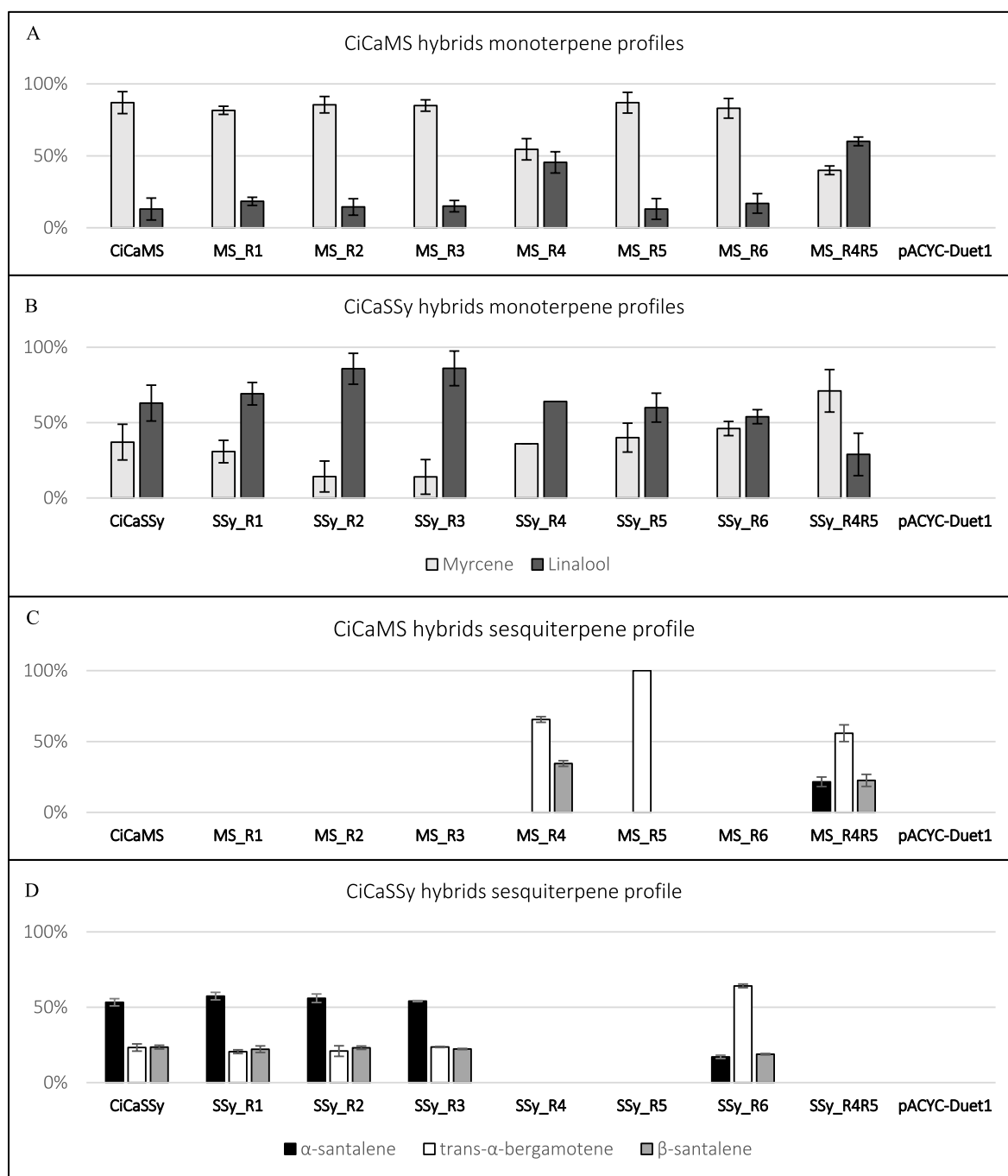
As a next step, the roles of 10 individual positions in the amino acid

sequence for producing santalenes were investigated, by exchanging them between CiCaSSy and CiCaMS (Table S3). In addition to the profile, total sesquiterpene production in dodecane was analysed for the most relevant mutants.

The residues were grouped, based on their position in respect to the active site cavity. From R4, residues 267, 291 and 294 appear to belong to the active site (Fig. 3C). These residues were substituted in CiCaMS and CiCaSSy, obtaining MS\_S267N, MS\_L291I, MS\_F294M from CiCaMS and SSy\_N267S, SSy\_I291L and SSy\_M294F from CiCaSSy. When testing the product profile of these mutant enzymes only one residue appeared to be responsible for sesquiTPS activity. Mutant MS\_F294M showed low but well detectable production of trans- $\alpha$ -bergamotene. Conversely, complementary mutant SSy\_M294F had lost the ability to produce sesquiterpenes in this system. Mutation of residues N267 and I291 each resulted in a change of the product ratio in CiCaSSy (Fig. 6A–B), but no major change in total sesquiterpene production was observed. The more distant residues in R4, 273 and 308, were also probed for their role in terpene synthesis, by testing mutants MS\_G273A and MS\_E308D for CiCaMS and SSy\_A273G and SSy\_D308E for CiCaSSy. Among these, mutant SSy\_A273G displayed an altered sesquiterpene profile and lower productivity, compared to CiCaSSy. SSy\_D308E showed a product profile which was comparable to the wild type CiCaSSy.

The same approach was used for region R5, where positions 401, 403 and 404 participate in the active site, and positions 415 and 419 appear to be located further away, near the bottom of the cavity (Fig. 3C). Among the CiCaMS mutants, production of trans- $\alpha$ -bergamotene was observed for both MS\_L403F and MS\_L404V (Fig. 6C). Apparently, either of these mutations in CiCaMS is sufficient to confer sesquiTPS activity, albeit that the product accumulation was lower than observed for CiCaSSy and hybrid MS\_R5. Conversely, substitution V404L in CiCaSSy resulted in a complete loss of sesquiTPS activity. CiCaSSy mutants A401V and F403L displayed altered product ratios compared to wild-type CiCaSSy, with trans- $\alpha$ -bergamotene being the dominant product (Fig. 6D). Mutations in the two residues further away from the active site did not alter the product spectrum of CiCaSSy (SSy\_Q415H, SSy\_E419A), nor did they confer sesquiterpene synthase activity on CiCaMS (MS\_H415Q, MS\_A419E).

Thus, the substitution analysis of region R4 and R5 indicates that three single amino acid positions are crucial to introduce sesquiTPS activity in CiCaMS: substitutions F294M, L403F and L404V each result in sesquiTPS activity of the monoTPS. In CiCaSSy, substitutions M294F



**Figure 5.** (A–B) Monoterpene profiles of relevant hybrids derived from (A) CiCaMS and (B) CiCaSSy. Activity was tested in an *in vitro* system using cell-free extracts and GDP as substrate. (C–D) Profile of produced sesquiterpenes by enzyme hybrids in an *E. coli* fermentation system. (C) Sesquiterpene production profiles mediated by CaCaMS and its derived hybrids; (D) sesquiterpene production profiles mediated by CiCaSSy and its derived hybrids. (A–D) Values on the Y axis express the relative ratio of each compound, relative to the total sesquiterpenes. Each variant was tested in three independent experiments. Error bars indicate the standard deviation.

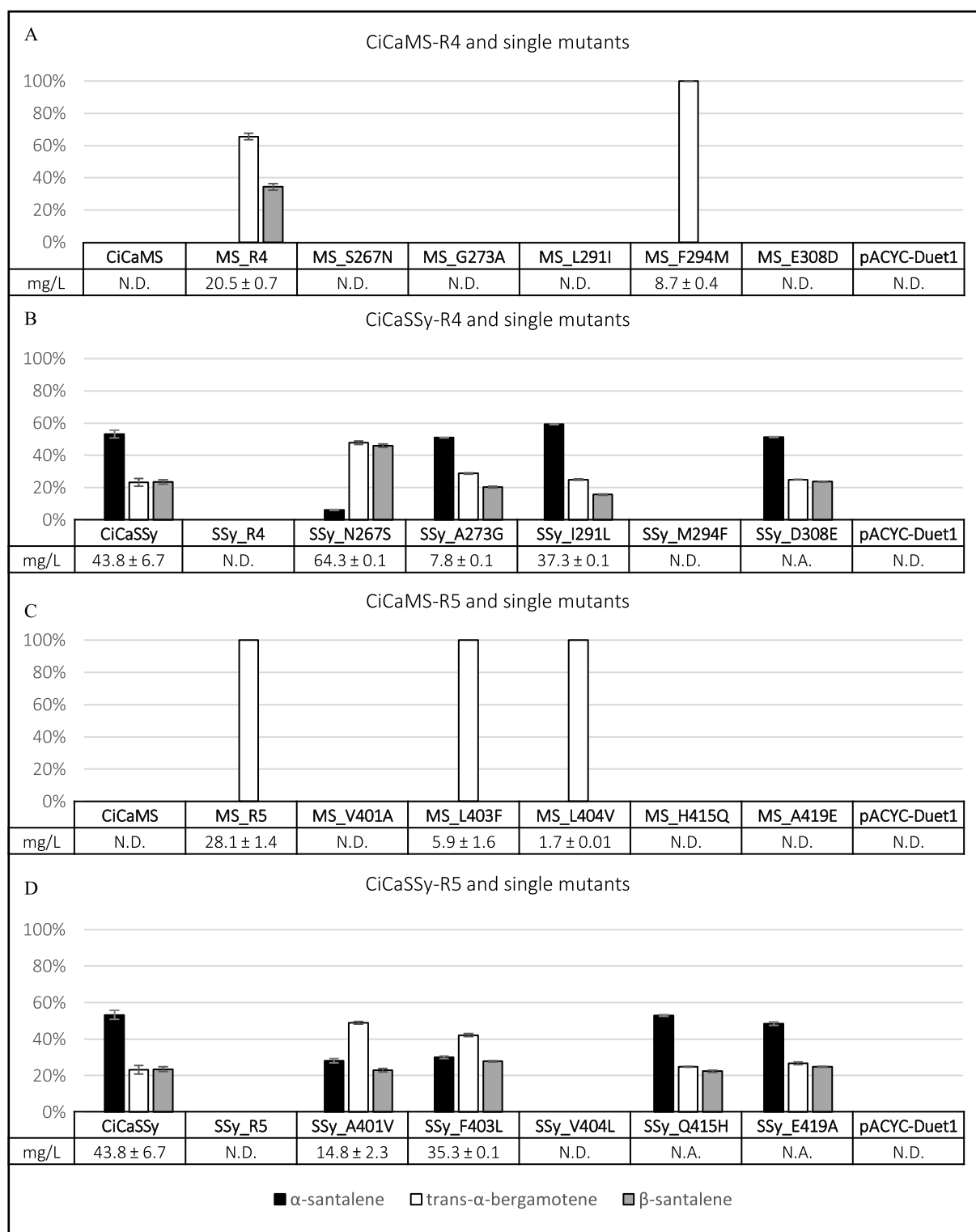
and V404L result in a complete loss of sesquiTPS activity.

Although R6 was not observed to be crucial for sesquiterpene production, its exchange resulted in a significant variation in product ratio (Fig. 7). Among the R6 variant residues, residue D442 maps close to the active site of CiCaSSy and is part of its NSE/DTE motif (Fig. 3B–C). Two mutations, MS\_N442D and SSy\_D442N, were tested in variant MS\_R4R5 and in CiCaSSy, respectively. As shown in Fig. 7A, hybrid MS\_R4R5 with the additional single mutation N442D recovers the product ratio of the wild type CiCaSSy. Conversely, SSy\_D442 N results in the same product ratio as SSy\_R6 and MS\_R4R5 (Fig. 7B). This confirms residue 442 as a

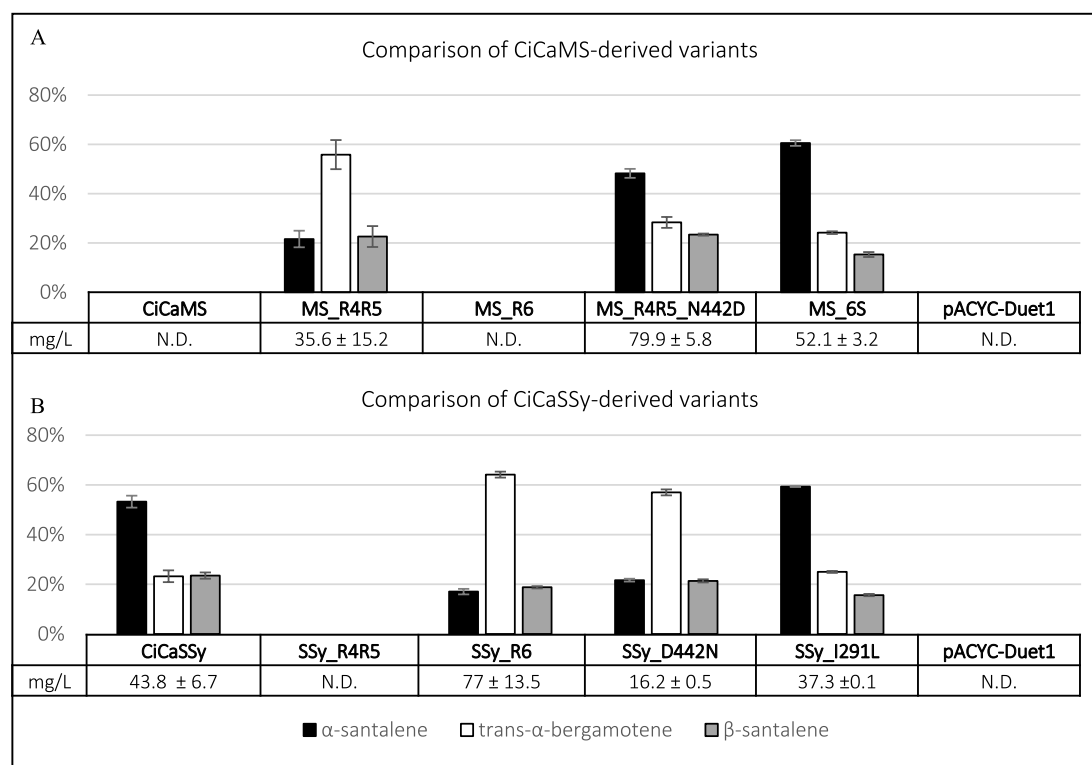
major determinant of product ratio in CiCaSSy.

Based on these results, we hypothesised that the combination of the six residues identified above (267, 294, 401, 403, 404, 442) would be sufficient to effectively establish the CiCaSSy sesquiTPS profile in CiCaMS. To substantiate this, we generated a CiCaMS-derived mutant carrying all these six amino acid substitutions. The obtained variant, referred to as MS\_6S, was at least as active as CiCaSSy in producing sesquiterpenes and showed a product profile very similar to wild type CiCaSSy (Fig. S9), with the presence of all three compounds and α-santalene as major peaks (Fig. 7B).





**Fig. 6.** Profile of produced sesquiterpenes by amino-acid mutants in an *E. coli* fermentation system. (A) sesquiterpene production profiles mediated by CiCaMS R4 hybrid and single mutants; (B) sesquiterpene production profiles mediated by CiCaSSy R4 hybrid and single mutants. (C) sesquiterpene production profiles mediated by CiCaMS R5 hybrid and single mutants; (D) sesquiterpene production profiles mediated by CiCaSSy R5 hybrid single mutants. Activity of WT enzymes has been included for comparison. Values on the Y axis express the relative ratio of each compound, relative to the total sesquiterpenes produced. Each variant was tested in three independent experiments. Error bars indicate the standard deviation, tables report the calculated mg/L of total sesquiterpenes produced. In the tables, N.D. stands for “not detected” while N.A. for “not analysed”.



**Fig. 7.** Profile of produced sesquiterpenes by (A) CiCaMS and (B) CiCaSSy variants in an *E. coli* fermentation system. (A) sesquiterpene production profiles mediated by CiCaMS double hybrid MS\_R4R5 in comparison with the single mutant MS\_R4R5\_N442D. (B) sesquiterpene production profiles mediated by SSy\_R6 hybrid and R6 single mutant (SSy\_D442N). In graphic (A) is also reported MS\_6S, in comparison with SSy\_I291L in (B). Error bars indicate standard deviation, tables report the calculated mg/L of total sesquiterpene produced. In the tables, N.D.: not detected; N.A.: not analysed.

#### 4. Discussion

In our study we characterized a novel santalene synthase from *C. camphora* (CiCaSSy) which shows low similarity with the previously characterized santalene synthases from *Santalum* spp. (~38% identity [20]). We also identified a closely related monoTPS (CiCaMS) which does not show any sesquiterpene activity despite differing from CiCaSSy in only 22 out of 553 (95%) amino acids. Among these, three residue changes (M294F, L403F and V404L) were each able to convert the monoTPS into a sesquiTPS. Three more substitutions (S267N, V401A and N442D) appear to be involved in defining the product profile of CiCaSSy. Thus, six residues define the specific product properties of CiCaSSy, relative to its monoTPS counterpart. To examine the amino acid differences between these two enzymes in a functional context, we used a structural model of CiCaSSy, depicted in Fig. 3B.

Two of the most important residues addressed in this study are at position 294 and 404. Both in CiCaSSy and in CiCaMS, the identity of the residues in these positions determine sesquiTPS activity. In the structure model, the sidechains of these residues point into the active site cavity, although the precise topology of the side chain cannot be accurately inferred from the structure model. For residue 294, the size properties of the Phe side chain as found in CiCaMS may hinder entry of FDP into the active site pocket, thus preventing sesquiTPS activity. The role of this region in TPS function has been earlier addressed by Kampranis et al. (2007) [49], who showed that in the 1,8-cineole monoTPS from *Salvia fruticosa*, a mutation of an Asn to Ala, in a position corresponding to 291 in CiCaSSy, allowed for the enlargement of the active site cavity to accommodate the bulkier FDP substrate and induce sesquiterpene production. Both position 291 and 294 have differing amino acids in the two *Cinnamomum* synthases, but only 294 appears to affect sesquiterpene production. An alignment of relevant synthases, highlighting the residues of interest, is presented in Fig. S7.

Residue 404, together with residues 401 and 403, lies around the

kink in the G2 helix, which has been studied in many different contexts as being crucial for product specificity in TPSs [49–52]. These previous studies reveal that mutations in these positions can lead to changes in the product specificity, consistently with our observations that substitutions in positions 401 and 403 have a strong impact on the product profile of CiCaSSy. However, a stronger effect is observed for position 404, which induces sesquiTPS activity in CiCaMS (MS\_L404V) and disrupts completely the sesquiTPS activity in CiCaSSy (SSy\_V404L). The results presented here indicate that conversion of CiCaMS into a bergamotene synthase can be mediated by a variety of mutations which affect the shape of the active site cavity.

More subtle changes result from mutations in positions 267, 291 and 442, which predominantly have an effect on the product profile of the sesquiTPS, leading to altered ratios of trans-α-bergamotene, α- and β-santalene. Position 267 (Ser in CiCaMS and Asn in CiCaSSy; Fig. 6B) has been implicated in the second cyclization required to produce bicyclic monoterpenes such as α-pinene [53], corresponding to the third cyclization for the production of tricyclic sesquiterpenes. Thus, this functional activity may explain the difference in product ratio of the bicyclic sesquiterpenes (trans-α-bergamotene and β-santalene) compared to the tricyclic α-santalene seen in SSy\_N267S (Fig. 6B). Residue 442 in R6 forms part of the catalytic NSE/DTE motif, which acts as the second Mg<sup>2+</sup> binding motif in TPSs [54]. Our previous research [13] showed that this position is predominantly (65%) an Asp among 250 characterized plant sesquiTPSs. Possibly, the prominent role of this residue in determining the product profile of CiCaSSy is related to its involvement in orienting the magnesium ion.

The residues which determine myrcene synthase activity in CiCaMS are mostly located in region R4: when R4 from CiCaSSy is introduced in CiCaMS, the product ratio is affected at the expense of myrcene, while in other hybrids, the ratio between myrcene and linalool is maintained. Interestingly, the myrcene-dominated profile is maintained in MS\_R5, which is yet also able to produce the sesquiterpene trans-alpha

bergamotene. The role of individual residues in R4 and R5 for myrcene biosynthesis was not further investigated, as the main focus of this work is the identification of residues which determine the multiproduct sesquiterpene synthase profile.

The fact that CiCaSSy and CiCaMS are so closely related raises the question whether the common ancestor enzyme was a sesquiTPS or a monoTPS. Our results do not provide clear answers to this question. The highest sequence similarity of both enzymes was found with a terpineol synthase, suggesting that the santalene synthase CiCaSSy could have evolved from a more common monoTPS present in *Cinnamomum* spp. On the other hand, the ability of CiCaSSy to produce a suite of bicyclic and tricyclic sesquiterpenes, which require a complex cascade of proton migrations in the sesquiterpene, as opposed to the presumed more simple linear monoterpene myrcene, which requires hardly any proton transfer, could lead to consider CiCaMS as a sort of loss-of-function mutant of CiCaSSy [38,55,56]. This view would also be in agreement with the absence of a clear plastid transit peptide in both CiCaMS and CiCaSSy. However, it should be noted that CiCaMS is a functional monoterpene synthase, since it does mediate myrcene formation when expressed in *N. benthamiana*. Moreover, the chromatogram of the root tissue (Fig. 1) shows a myrcene peak consistent with the activity observed for CiCaMS. Thus, different hypotheses into the evolutionary order of myrcene and santalene evolution find support.

TPSs with a mixed mono/sesquiterpene product profile have been described before. Several studies describe the ability of few sesquiTPS to also behave as monoTPSs in the presence of GDP [57]. The santalene synthases and the bisabolene synthases from *Santalum* spp. have also been observed to produce linalool, geraniol and terpineol when supplied with GDP [20]. Other examples are the trans- $\alpha$ -bergamotene synthase from *Lavandula angustifolia* [58] and the  $\alpha$ -bisabolene synthase from *Abies grandis* [59]. Interestingly, all these enzymes result in products which are part of a specific subclass of sesquiterpenes, which derive from the bisabolyl cation [60]. One could hypothesize that sesquiterpenes derived from the bisabolyl cation, which present a cyclized “head” and a uncyclized “tail”, can be produced by synthases which are closely related to monoTPSs, and may have evolved from them. From this perspective, santalenes, bergamotenes and bisabolenes can also be seen as cyclized monoterpenes with an isopentenyl extension. As mentioned above, a change in residue 294, 403, or 404 seems to be sufficient to change the shape of the active site pocket of CiCaMS and to allow the accommodation of the larger FDP substrate. Thus, this hypothesis is sustained both by the ability of CiCaSSy to produce monoterpenes *in vitro* and, more importantly, by the demonstration that several single residue substitutions in the active site of monoTPS CiCaMS each are sufficient to trigger the production of sesquiterpene trans- $\alpha$ -bergamotene.

## 5. Conclusions

With this study we characterized two novel TPSs, one monoTPSs and one sesquiTPS from *C. camphora*. Residues essential for the conversion of the monoTPS into a sesquiTPS were identified and we effectively succeeded into converting CiCaMS in a santalene synthase showing the same product profile as CiCaSSy, by substituting only six residues. This work provides new insights into the function of specific residues and their role in the catalytic site of TPSs, contributing to a better understanding of this class of enzymes.

## Declaration of competing interest

Authors JB, AvH and DB are listed as inventors in a pending patent application (WO2018/160066; applicant, Isobionics) encompassing information presented here.

## Acknowledgements

This work is part of the research programme Novel Enzymes for Flavour and Fragrance with project number TTW 15043 which is financed by the Netherlands Organisation for Scientific Research (NWO), grant TTW 15043. Celina Vossen is acknowledged for her support with quantitative analysis of sesquiterpenes.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.abb.2020.108647>.

## Database

Sequences of the parental enzymes are available in GenBank with the IDs: MN756611 (CiCaMS) and MN756612 (CiCaSSy)

## Enzymes

CiCaSSy, santalene synthase (EC4.2.3.81; EC4.2.3.82; EC4.2.3.83) from *Cinnamomum camphora*; CiCaMS, myrcene synthase (EC4.2.3.15) from *Cinnamomum camphora*.

## References

- [1] M. Balandrin, J. Klocke, E. Wurtele, W. Bollinger, Natural plant chemicals: sources of industrial and medicinal materials, *Science* 228 (1985) 1154–1160, <https://doi.org/10.1126/science.3890182>.
- [2] L. Alves Gomes Albertti, T.L. Delatte, K. Souza de Farias, A. Galdi Boaretto, F. Verstappen, A. van Houwelingen, K. Cankar, C.A. Carollo, H.J. Bouwmeester, J. Beekwilder, Identification of the bisabolol synthase in the endangered candeia tree (*Eremanthus erythropappus* (DC) McLeisch), *Front. Plant Sci.* 9 (2018), <https://doi.org/10.3389/fpls.2018.01340>.
- [3] D. Bhaskar, S. Viswanath, S. Purushothaman, Sandal (*Santalum album* L.) conservation in Southern India: a review of policies and their impacts, *J. Trop. Agric.* 48 (2010) 1–10, accessed, [https://www.researchgate.net/publication/266268084\\_Sandal\\_Santalum\\_album\\_L\\_conservation\\_in\\_southern\\_India\\_A\\_review\\_of\\_policies\\_and\\_their\\_impacts](https://www.researchgate.net/publication/266268084_Sandal_Santalum_album_L_conservation_in_southern_India_A_review_of_policies_and_their_impacts). (Accessed 21 October 2019).
- [4] J.-F. Butaud, J.-P. Bianchini, J.-M. Bouvet, V. Gaydou, É. Lhuillier, P. Raharivelomanana, F. Rives, Sandalwood, Current State of Knowledge and Implications for Conservation and Enhancement, 2014 accessed, [https://www.researchgate.net/publication/282507135\\_Sandalwood\\_current\\_state\\_of\\_knowledge\\_and\\_implications\\_for\\_conservation\\_and\\_enhancement](https://www.researchgate.net/publication/282507135_Sandalwood_current_state_of_knowledge_and_implications_for_conservation_and_enhancement). (Accessed 21 October 2019).
- [5] A. Smit, Biosynthesis of isoprenoids via mevalonate in Archaea: the lost pathway, *Genome Res.* 10 (2000) 1468–1484, <https://doi.org/10.1101/gr.145600>.
- [6] J. Degenhardt, T.G. Köllner, J. Gershenzon, Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants, *Phytochemistry* 70 (2009) 1621–1637, <https://doi.org/10.1016/j.phytochem.2009.07.030>.
- [7] A. Aharoni, M.A. Jongsma, H.J. Bouwmeester, Volatile science? Metabolic engineering of terpenoids in plants, *Trends Plant Sci.* 10 (2005) 594–602, <https://doi.org/10.1016/j.tplants.2005.10.005>.
- [8] P.-H. Liang, T.-P. Ko, A.H.J. Wang, Structure, mechanism and function of prenyltransferases, *Eur. J. Biochem.* 269 (2002) 3339–3354, <https://doi.org/10.1046/j.1432-1033.2002.03014.x>.
- [9] B.M. Lange, T. Rujan, W. Martin, R. Croteau, Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes, *Proc. Natl. Acad. Sci. Unit. States Am.* 97 (2000) 13172–13177, <https://doi.org/10.1073/pnas.240454797>.
- [10] V.J.J. Martin, D.J. Pitera, S.T. Withers, J.D. Newman, J.D. Keasling, Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids, *Nat. Biotechnol.* 21 (2003) 796–802, <https://doi.org/10.1038/nbt833>.
- [11] I. Pateraki, A.M. Heskes, B. Hamberger, Cytochromes p450 for terpene functionalisation and metabolic engineering, *Adv. Biochem. Eng. Biotechnol.* 148 (2015) 107–139, [https://doi.org/10.1007/10.2014\\_301](https://doi.org/10.1007/10.2014_301).
- [12] D.W. Christianson, Structural and chemical biology of terpenoid cyclases, *Chem. Rev.* 117 (2017) 11570–11648, <https://doi.org/10.1021/acs.chemrev.7b00287>.
- [13] J. Durairaj, A. Di Girolamo, H.J. Bouwmeester, D. de Ridder, J. Beekwilder, A. D. van Dijk, An analysis of characterized plant sesquiterpene synthases, *Phytochemistry* 158 (2019) 157–165, <https://doi.org/10.1016/j.phytochem.2018.10.020>.
- [14] J. Beekwilder, A. van Houwelingen, K. Cankar, A.D.J. van Dijk, R.M. de Jong, G. Stoopen, H. Bouwmeester, J. Achkar, T. Sonke, D. Bosch, Valencene synthase from the heartwood of Nootka cypress (*Callitropsis nootkatensis*) for biotechnological production of valencene, *Plant Biotechnol. J.* 12 (2014) 174–182, <https://doi.org/10.1111/pbi.12124>.

- [15] J.A. Teixeira da Silva, M.M. Kher, D. Soner, T. Page, X. Zhang, M. Nataraj, G. Ma, Sandalwood: basic biology, tissue culture, and genetic transformation, *Planta* 243 (2016) 847–887, <https://doi.org/10.1007/s00425-015-2452-8>.
- [16] A.N. Arunkumar, A. Dhyani, G. Joshy, *Santalum Album* IUCN Red List Assessment, 2019, <https://doi.org/10.2305/IUCN.UK.2019-1>.
- [17] J.M. Celedon, A. Chiang, M.M.S. Yuen, M.L. Diaz-Chavez, L.L. Madilao, P. M. Finnegan, E.L. Barbour, J. Bohlmann, Heartwood-specific transcriptome and metabolite signatures of tropical sandalwood (*Santalum album*) reveal the final step of (Z)-santalol fragrance biosynthesis, *Plant J.* 86 (2016) 289–299, <https://doi.org/10.1111/tj.13162>.
- [18] M.L. Diaz-Chavez, J. Moniodis, L.L. Madilao, S. Jancsik, C.I. Keeling, E.L. Barbour, E.L. Ghisalberti, J.A. Plummer, C.G. Jones, J. Bohlmann, Biosynthesis of sandalwood oil: *Santalum album* CYP76F cytochromes P450 produce santalols and bergamotol, *PLoS One* 8 (2013), e75053, <https://doi.org/10.1371/journal.pone.0075053>.
- [19] A.A. Birkbeck, The synthesis of fragrant natural products from *Santalum album* L.: (+)-(Z)- $\alpha$ -Santalol and (–)-(Z)- $\beta$ -Santalol, *Chim. Int. J. Chem.* 71 (2017) 823–835, <https://doi.org/10.2533/chimia.2017.823>.
- [20] C.G. Jones, J. Moniodis, K.G. Zulak, A. Scaffidi, J.A. Plummer, E.L. Ghisalberti, E. L. Barbour, J. Bohlmann, Sandalwood fragrance biosynthesis involves sesquiterpene synthases of both the terpene synthase (TPS)-a and TPS-b families, including santalene synthases, *J. Biol. Chem.* 286 (2011) 17445–17454, <https://doi.org/10.1074/jbc.M111.231787>.
- [21] E. Jongedijk, K. Cankar, J. Ranzijn, S. van der Krol, H. Bouwmeester, J. Beekwilder, Capturing of the monoterpene olefin limonene produced in *Saccharomyces cerevisiae*, *Yeast* 32 (2015) 159–171, <https://doi.org/10.1002/yea.3038>.
- [22] K. Cankar, E. Jongedijk, M. Klompmaier, T. Majdic, R. Mumm, H. Bouwmeester, D. Bosch, J. Beekwilder, (+)-Valencene production in *Nicotiana benthamiana* is increased by down-regulation of competing pathways, *Biotechnol. J.* 10 (2015) 180–189, <https://doi.org/10.1002/biot.201400288>.
- [23] J. Quan, J. Tian, Circular polymerase extension cloning of complex gene libraries and pathways, *PLoS One* 4 (2009), <https://doi.org/10.1371/journal.pone.0006441>.
- [24] Y. Xia, W. Chu, Q. Qi, L. Xun, New insights into the QuikChange™ process guide the use of Phusion DNA polymerase for site-directed mutagenesis, *Nucleic Acids Res.* 43 (2015) e12, <https://doi.org/10.1093/nar/gku1189>.
- [25] P.P. Peralta-Yahya, M. Ouellet, R. Chan, A. Mukhopadhyay, J.D. Keasling, T.S. Lee, Identification and microbial production of a terpene-based advanced biofuel, *Nat. Commun.* 2 (2011) 483, <https://doi.org/10.1038/ncomms1494>.
- [26] J.S. Reece-Hoyes, A.J.M. Walhout, Gateway recombinational cloning, *Cold Spring Harb. Protoc.* (2018), top094912, <https://doi.org/10.1101/pdb.top094912>.
- [27] F.A. van Engelen, J.W. Molthoff, A.J. Conner, J.-P. Nap, A. Pereira, W.J. Stiekema, pBINPLUS: an improved plant transformation vector based on pBIN19, *Transgenic Res.* 4 (1995) 288–290, <https://doi.org/10.1007/BF01969123>.
- [28] J. Sambrook, D.W. Russell, SDS-polyacrylamide gel Electrophoresis of proteins, *Cold Spring Harb. Protoc.* (2006), <https://doi.org/10.1101/pdb.prot4540> pdb. prot4540.
- [29] N. Eswar, B. Webb, M.A. Marti-Renom, M.S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper, A. Sali, Comparative protein structure modeling using modeller, *Curr. Protoc. Bioinforma.* 15 (2006), <https://doi.org/10.1002/0471250953.bio506s15.5.6.1-5.6.30>.
- [30] B. Boeckmann, A. Bairoch, R. Apweiler, M.C. Blatter, A. Estreicher, E. Gasteiger, M. J. Martin, K. Michoud, C. O'Donovan, I. Phan, S. Pilboud, M. Schneider, The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, *Nucleic Acids Res.* (2003), <https://doi.org/10.1093/nar/gkg095>.
- [31] W. Li, A. Godzik, Cd-hit, A fast program for clustering and comparing large sets of protein or nucleotide sequences, *Bioinformatics* (2006), <https://doi.org/10.1093/bioinformatics/btl158>.
- [32] F. Sievers, A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J.D. Thompson, D.G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol.* 7 (2011), <https://doi.org/10.1038/msb.2011.75>.
- [33] R.D. Finn, A. Bateman, J. Clements, P. Coghill, R.Y. Eberhardt, S.R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E.L.L. Sonnhammer, J. Tate, M. Punta, Pfam: the protein families database, *Nucleic Acids Res.* 42 (2014), <https://doi.org/10.1093/nar/gkt1223>.
- [34] S. Capella-Gutiérrez, J.M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, *Bioinformatics* (2009), <https://doi.org/10.1093/bioinformatics/btp348>.
- [35] J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: reconstruction, analysis, and visualization of phylogenomic data, *Mol. Biol. Evol.* 33 (2016) 1635–1638, <https://doi.org/10.1093/molbev/msw046>.
- [36] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, *Bioinformatics* 30 (2014) 1312–1313, <https://doi.org/10.1093/bioinformatics/btu033>.
- [37] I. Letunic, P. Bork, Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation, *Bioinformatics* 23 (2007) 127–128, <https://doi.org/10.1093/bioinformatics/btl529>.
- [38] F. Chen, D. Tholl, J. Bohlmann, E. Pichersky, 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 (2011) 212–229, <https://doi.org/10.1111/j.1365-3113.2011.04520.x>.
- [39] B.J. Stubbs, A. Specht, D. Brushett, The essential oil of *Cinnamomum camphora* (L.) nees and Eberm — variation in oil composition throughout the tree in two chemotypes from Eastern Australia, *J. Essent. Oil Res.* 16 (2004) 200–205, <https://doi.org/10.1080/10412905.2004.9698697>.
- [40] Y. Pellissier, C. Marion, S. Prunac, J.-M. Bessière, Volatile Components of Leaves, Stems and Bark of *Cinnamomum camphora* Nees et Ebermaier, *J. Essent. Oil Res.* 7 (1995) 313–315, <https://doi.org/10.1080/10412905.1995.9698525>.
- [41] C.D. Frizzo, A.C. Santos, N. Paroul, L.A. Serafini, E. Dellacassa, D. Lorenzo, P. Moyna, Essential oils of camphor tree (*Cinnamomum camphora* Nees ^ Eberm) cultivated in Southern Brazil, *Braz. Arch. Biol. Technol.* 43 (2000) 313–316, <https://doi.org/10.1590/S1516-89132000000300011>.
- [42] J.A. Pino, V. Fuentes, Leaf oil of *Cinnamomum camphora* (L.) J. Presl. From Cuba, *J. Essent. Oil Res.* 10 (1998) 531–532, <https://doi.org/10.1080/10412905.1998.9700962>.
- [43] C. Chen, Y. Zheng, Y. Zhong, Y. Wu, Z. Li, L.-A. Xu, M. Xu, Transcriptome analysis and identification of genes related to terpenoid biosynthesis in *Cinnamomum camphora*, *BMC Genom.* 19 (2018) 550, <https://doi.org/10.1186/s12864-018-4941-1>.
- [44] R.P. Adams, Identification of Essential Oil Components by Gas Chromatography/Quadrupole Mass Spectroscopy, fourth ed., Allured Pub Corp, 1995 accessed, [http://www.researchgate.net/publication/283650275\\_Identification\\_of\\_Essential\\_Oil\\_Components\\_by\\_Gas\\_ChromatographyQuadrupole\\_Mass\\_Spectroscopy](http://www.researchgate.net/publication/283650275_Identification_of_Essential_Oil_Components_by_Gas_ChromatographyQuadrupole_Mass_Spectroscopy). (Accessed 22 October 2019).
- [45] E.W. Sayers, R. Agarwala, E.E. Bolton, J.R. Brister, K. Canese, K. Clark, R. Connor, N. Fiorini, K. Funk, T. Hefferon, J.B. Holmes, S. Kim, A. Kimchi, P.A. Kitts, S. Lathrop, Z. Lu, T.L. Madden, A. Marchler-Bauer, L. Phan, V.A. Schneider, C. L. Schoch, K.D. Pruitt, J. Ostell, Database resources of the national center for Biotechnology information, *Nucleic Acids Res.* 47 (2018) 23–28, <https://doi.org/10.1093/nar/gky1069>.
- [46] S.-M. Chaw, Y.-C. Liu, Y.-W. Wu, H.-Y. Wang, C.-Y.I. Lin, C.-S. Wu, H.-M. Ke, L.-Y. Chang, C.-Y. Hsu, H.-T. Yang, E. Sudianto, M.-H. Hsu, K.-P. Wu, L.-N. Wang, J. H. Leebens-Mack, I.J. Tsai, Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution, *Native Plants* 5 (2019) 63–73, <https://doi.org/10.1038/s41477-018-0337-0>.
- [47] S. Lee, J. Chappell, Biochemical and genomic characterization of terpene synthases in *Magnolia grandiflora*, *Plant Physiol.* 147 (2008) 1017–1033, <https://doi.org/10.1104/pp.108.115824>.
- [48] D.M. Martin, J. Bohlmann, Identification of Vitis vinifera (–)- $\alpha$ -terpineol synthase by in silico screening of full-length cDNA ESTs and functional characterization of recombinant terpene synthase, *Phytochemistry* 65 (2004) 1223–1229, <https://doi.org/10.1016/j.phytochem.2004.03.018>.
- [49] S.C. Kampranis, D. Ioannidis, A. Purvis, W. Mahrez, E. Ninga, N.A. Katerelos, S. Anssour, J.M. Dunwell, J. Degenhardt, A.M. Makris, P.W. Goodenough, C. B. Johnsona, Rational conversion of substrate and product specificity in a Salvia monoterpene synthase: structural insights into the evolution of terpene synthase function, *Plant Cell* 19 (2007), <https://doi.org/10.1105/tpc.106.047779>, 1994–2005.
- [50] M. Loizzi, V. González, D.J. Miller, R.K. Allemann, Nucleophilic water capture or proton loss: single amino acid switch converts  $\delta$ -cadinene synthase into germacadien-4-ol synthase, *Chembiochem* 19 (2018) 100–105, <https://doi.org/10.1002/cbic.201700531>.
- [51] T.G. Köllner, C. Schnee, J. Gershenzon, J. Degenhardt, The variability of sesquiterpene emitted from two Zea mays cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes, *Plant Cell* 16 (2004) 1115–1131, <https://doi.org/10.1105/tpc.019877>.
- [52] S. Garms, F. Chen, W. Boland, J. Gershenzon, T.G. Köllner, A single amino acid determines the site of deprotonation in the active center of sesquiterpene synthases SbTPS1 and SbTPS2 from Sorghum bicolor, *Phytochemistry* 75 (2012) 6–13, <https://doi.org/10.1016/j.phytochem.2011.12.009>.
- [53] L.-J. Hsu, F.-H. Chu, Plasticity residues involved in secondary cyclization of terpene synthesis in Taiwan cryptomerioides, *Tree Genet. Genomes* 11 (2015) 796, <https://doi.org/10.1007/s11295-014-0796-z>.
- [54] D.W. Christianson, Structural biology and chemistry of the terpenoid cyclases, *Chem. Rev.* 106 (2006) 3412–3442, <https://doi.org/10.1021/cr050286w>.
- [55] J. Bohlmann, G. Meyer-Gauen, R. Croteau, Plant terpenoid synthases: Molecular biology and phylogenetic analysis, *Proc. Natl. Acad. Sci. Unit. States Am.* 95 (1998) 4126–4133, <https://doi.org/10.1073/pnas.95.8.4126>.
- [56] P.S. Karunanithi, P. Zerbe, Terpene synthases as metabolic gatekeepers in the evolution of plant terpenoid chemical diversity, *Front. Plant Sci.* 10 (2019), <https://doi.org/10.3389/fpls.2019.01166>.
- [57] L. Pazouki, Ü. Niinemets, Multi-substrate terpene synthases: their occurrence and physiological significance, *Front. Plant Sci.* 7 (2016), <https://doi.org/10.3389/fpls.2016.01019>.
- [58] C. Landmann, B. Fink, M. Festner, M. Dregus, K.-H. Engel, W. Schwab, Cloning and functional characterization of three terpene synthases from lavender (*Lavandula angustifolia*), *Arch. Biochem. Biophys.* 465 (2007) 417–429, <https://doi.org/10.1016/j.abb.2007.06.011>.
- [59] J. Bohlmann, J. Crock, R. Jetter, R. Croteau, Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)- $\alpha$ -bisabolene synthase from grand fir (*Abies grandis*), *Proc. Natl. Acad. Sci. Unit. States Am.* 95 (1998) 6756–6761, <https://doi.org/10.1073/pnas.95.12.6756>.
- [60] IUBMB, bisabolene biosynthesis. <https://www.qmul.ac.uk/sbcs/iubmb/enzyme/reaction/terp/bisabol.html>, 2011. (Accessed 9 November 2019).